AFRL-HE-WP-TR-2002-0146

UNITED STATES AIR FORCE RESEARCH LABORATORY

Measurement and Modeling of Human Performance Under Differing G Conditions

Robert D. O'Donnell
Samuel L. Moise

NTI, INC.
4130 Linden Ave., Suite 235
Dayton OH 45432

February 2001

Final Report for the Period May 2000 to February 2001

20021129 079

Approved for public release; distribution is unlimited.

Human Effectiveness Directorate
Biodynamics and Protection Division
Biodynamics and Acceleration Branch
2800 Q Street
Wright-Patterson AFB OH 45433-7947

NOTICES

DISCLAIMER

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2002-0146

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE DIRECTOR

F. Wesley Baumgardner

F. WESLEY BAUMGARDNER, PhD
Chief, Biodynamics and Protection Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
February 2001	Final Report	May 2000 - February 2001

4. TITLE AND SUBTITLE
Measurement and Modeling of Human Performance Under Differing G Conditions

5a. CONTRACT NUMBER F33615-00-M-6032
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER 65502F

6. AUTHOR(S)
Robert D. O'Donnell
Samuel L. Moise

5d. PROJECT NUMBER 3005
5e. TASK NUMBER HP
5f. WORK UNIT NUMBER 0A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NTI, Inc.
4130 Linden Ave., Suite 235
Dayton OH 45432

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Human Effectiveness Directorate
Biodynamics and Protection Division
Biodynamics and Acceleration Branch
2800 Q Street, Bldg 824, Rm 206
Wright-Patterson AFB OH 45433-7947

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL/HE-WP-TR-2002-0146

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Report developed under SBIR contract for topic AF00-097. A significant need exists to develop field-usable techniques allowing a commander to estimate the performance effects of acceleration forces on the pilot-warfighter. This is required if countermeasures are to be evaluated, and if training and/or combat G-exposure standards are to be developed. In Phase I of this SBIR effort, NTI, Inc. developed a performance test methodology structured around a comprehensive model of human performance, and conceptualized a methodology that would translate performance data into operational military impact (OMI) estimates. Twelve test procedures were described, capturing essential flight performance capacities. This was called the "G-Performance Assessment Simulation System " (G-PASS). In addition, a new software product was also described - the "G-Tool to Optimize Performance" (G-TOP). This allows the user to input any G-profile and automatically generate performance predictions. These predictions are then entered into computer models of military operations to generate the required OMI in terms of any operationally meaningful measure of merit. These developments are described in the present report. Phase II of this effort will fully implement the concepts developed and described herein.

15. SUBJECT TERMS
SBIR Report, Performance Metrics, Acceleration, Centrifuge, Mission Modeling, Simulation, Part-task Simulation, Performance, Mission Effectiveness, Performance Optimization

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
UC	UC	UC	UL	55	Tamara L. Chelette / (937) 255-4096

THIS PAGE LEFT BLANK INTENTIONALLY

PREFACE

The work supporting this report was carried out under a Phase I Small Business Innovation Research (SBIR) contract from the USAF Air Force Research Laboratory. The contract technical monitor for this effort was Dr. Tamara Chelette who, along with Dr. William Albery provided invaluable technical advice and support. The pilot consultants who supplied input at many points in this effort were Mr. Robert Shaw and Mr. William Ercoline. Without their sound advice, this program could not have been carried out.

THIS PAGE LEFT BLANK INTENTIONALLY

TABLE OF CONTENTS

THIS PAGE LEFT BLANK INTENTIONALLY

INTRODUCTION

Flight in modern fighter aircraft intrinsically results in the body being subjected to extreme physiological challenges that have potentially life-threatening effects. Unusual and extreme vestibular stimulation, heat, and high-speed angular acceleration constitute only a few of these challenges. The field of aerospace medicine has long recognized the need to quantify and protect against adverse effects of such stresses on the pilot and on the operation of the aircraft. The Phase I SBIR effort on which this proposal is based was designed to address a particular aspect of that operational need.

A great deal of effort has been expended in centrifuges and other dynamic environments directed to ameliorating many of the physiological insults generated by high speed, high altitude flight. Although these studies have resulted in remarkable success in addressing the physical and physiological effects of such flight, there has been considerably less success in addressing more subtle issues of human cognitive performance limits in these environments (Von Gierke, McCloskey, and Albery, 1991: McCloskey, Tripp, Chelette, and Popper, 1992). In other words, there still is no precise way to quantify the potential performance decrement caused by physiological stress on the pilot, or even the degree to which performance may be degraded by the very techniques which are employed to protect the pilot from the physical threat.

There are several reasons why the measurement of human performance in such unusual environments is difficult. First, the stress environment itself (including the intrinsic variability of the pilot population) makes performance assessment difficult, leading to predictions that must be made from a limited sample of behavior. As the stress becomes more intense, with accompanying shortening of subject exposures, the researcher is forced to retreat to more and more basic (and indirect) measures of performance. This inevitably results in an increased use of laboratory measures that have a decreased meaning to the operational community. Although there have been creative attempts to bridge the gap between laboratory and "field", it is an unfortunate fact that no approach to performance measurement in high physiological-stress environments has provided data phrased in the operationally meaningful terms demanded by the system designer or field commander (see review by Perez, et al, 1987).

In Phase I of this effort, we recognized the need to develop a technique that would result in valid, reliable measures of the operationally meaningful military impact of the kinds and degrees of physiological stress expected in real-world combat environments. Our approach depended heavily on the idea that comprehensive models of human performance (e.g., Newell, 1990: Anderson, 1993) have evolved to the point that they could validly be used to augment experimental data and therefore to 'fill in the gaps' that the stressful environment imposes. Therefore, in Phase I, we have attempted to develop a process that utilizes all existing data on human performance in acceleration environments, but also uses powerful computer modeling techniques to project beyond these data to performance predictions in situations that can not be studied in any practical way.

The goal of the Phase I effort was "... the design of an integrated performance testing and modeling approach that will permit assessment of the "Operational Military Impact" (OMI) of altered G environments." In other words, our goal was to utilize existing acceleration literature to provide the warfighter with meaningful and defensible estimates of the effect of G forces on the pilot's performance.

In addressing this goal, NTI recognized the problems involved. As noted above, the experimental data involving performance assessment under high G forces is relatively sparse. It became clear that if data necessary to enter into human performance models were to be obtained, creative use of other sources would be necessary. Further, it was clear that the underlying model of human performance would require considerable attention. Many models exist, including those that hope to approach a "unifying theory" of human performance (Newell, 1990). However, it was not clear that these models were appropriate for the environment of interest. These and other obstacles had to be overcome if we were to achieve the overall goal.

Our approach to solving these problems basically addressed three tasks:

1) A survey of computer modeling technology with respect to human performance prediction in acceleration research.

2) Definition of performance assessment techniques appropriate for the centrifuge, and integration of these into the proposed model.

3) Most importantly, development of an innovative plan for integrating existing data, current human performance models, and other techniques to provide operationally useful measures to the warfighter on the performance effects of complex G-exposures.

ACTIVITIES AND RESULTS

Overview of the Effort

Over the nine months of this Phase I effort, NTI carried out the three tasks defined above, and these are described in some detail below. In summary, the final output of this program is the design and demonstration of a technology that will eventually permit operational planners and researchers to estimate the performance degradation or enhancement, if any, that will occur as a result of exposure to any combination of G forces. Although not all of the potential problems in conceptualizing this technology could be addressed in this initial effort, we believe that these can be solved, and that this new approach may constitute an entirely new solution to the problem. If successful, it could well point the direction that acceleration research (as well as other commercially related research and products) will take in the future.

The core of this new development is a predictive model of the human that can incorporate various stressors and result in: 1) an estimate of the performance impact on the human, and 2) an estimate of the "operational" impact (military or otherwise) of that performance impact. In other words, the system conceptualized and demonstrated in Phase I uses an underlying cognitive model (as well as actual data) to predict the operational performance outcome of any combination of G exposures. A simple illustration of how this system would work was produced using a CD-based "breadboard" that allows the user to input various G-profiles in order to see (hypothetical) examples of how these profiles would alter the person's ability in each of the skills contained in the underlying cognitive model. If fully developed and implemented we believe this system will provide a tool that will permit the field commander or operational planner to monitor/schedule pilot missions so as to optimize performance. For this reason, the tool is called the **"G-TOOL to OPTIMIZE PERFORMANCE" (G-TOP).**

Obviously, we recognize that the concept as developed in Phase I is innovative, and that it will require considerable careful and detailed development to become fully implemented. We also recognize that the model and approach described here have several points at which inferences from subject matter experts and other models must be used. We have tried to make these assumptions, data deficiencies, and other potential weaknesses explicit in this report, and to suggest ways in which these problems can be remedied.

DESCRIPTION OF THE HUMAN PERFORMANCE MODEL

One of the major goals of the Phase I effort was to survey the human performance literature, especially models of performance that have been developed. Although many models of performance exist, especially those addressing cognitive performance, none have specifically addressed the types of performance of interest in the acceleration environment. Considerable time was spent in reviewing a large number of these models during the early stages of this effort, with a view to determining whether any could be

adapted for use in that environment. A brief review of some of the more promising ones is presented below.

<u>Summary of Performance Models Considered</u>

A number of lesser-known models were considered and immediately rejected as being inappropriate for the present purpose. A sample of the models we felt were not suitable include:

1. 20-sim – Primarily for the simulation of dynamic systems such as electrical, mechanical and hydraulic systems.
2. ACSL Model – For complex, non-linear systems. Includes a visual programming language.
3. Arena – Creates animated models that represent virtually any system. This model is primarily focused on manufacturing.
4. AutoMod – This software differs significantly from other simulation systems because of its ability to deal with the physical elements of a system in physical (graphical) terms and the logical elements of a system in logical terms.
5. CPSim - A general-purpose simulation tool for writing discrete-event simulations in the C programming language.
6. Mesquite CSIM 17 - A process-oriented, general purpose simulation toolkit written with general C language functions.
7. Dymola - An object-oriented language and a program for modeling of large systems.
8. EASY5 - A graphical-user-interface based software used to model, analyze and design dynamic systems - systems described by differential, difference and algebraic equations.
9. Extend - A tool for creating discrete event and continuous simulations.
10. MCSim - A simulation package which allows you to design models and to perform Monte Carlo stochastic simulations, or Bayesian inference through Markov Chain Monte Carlo simulations.
11. ModelMaker - A simulation modeling package designed for scientists and engineers.
12. PowerSim – A modeling package geared primarily at compartmental models.
13. SIMPLE++ is the standard software for object-oriented, graphical and integrated modeling, simulation and animation of systems and business processes.
14. SIMSCRIPT II.5(R) is a powerful, free-form, English-like simulation language designed to simplify writing programs for simulation modeling.
15. SIMULINK is a tool for modeling, analyzing, and simulating an extraordinarily wide variety of physical and mathematical systems, including those with nonlinear elements and those which make use of continuous and discrete time.
16. SMPL is a general purpose discrete event simulation library written in C. SMPL is portable and uses an event scheduling (as opposed to activity or process) oriented view.

17. STELLA II - Provides a powerful modeling environment for the investigation of any time-dependent process (including discrete event simulation).
18. Taylor II - Discrete event simulation software.
19. Vensim - An integrated environment for developing, analyzing and packaging high quality dynamic feedback models.
20. VisSim - A visual block diagram language for nonlinear dynamic simulation.

We primarily focused on 12 well-known models. These models have a well established history of use, are based on well-accepted theoretical foundations, and could reasonably be implemented through performance tests of the type that could be used in the acceleration environment. These models are:

1. <u>Adaptive Control of Thought – Rational (ACT-R);</u>

The ACT-R unified theory of cognition attempts to develop a cognitive architecture that can perform in detail a full range of cognitive tasks. The architecture takes the form of a computer simulation that is capable of performing and learning from the same tasks worked on by human subjects in our laboratories.

2. <u>Cognition as a Network of Tasks (COGNET);</u>

COGNET (COGnition as a NEtwork of Tasks) is a theoretically based set of tools and techniques for performing cognitive task analyses and building models of human-computer interaction in real-time, multi-tasking environments. These models view cognitive processes as the operation of a specific computational mechanism on a set of symbols, which are themselves a representation of sensation, experience, and its abstraction.

3. <u>Distributed Operator Model Architecture (D-OMAR);</u>

D-OMAR is a discrete-event simulation environment ideally suited to meet the demands of modeling the command and control environment. D-OMAR provides specific capabilities to support the design, execution, and analysis of experiments. The actions and events that drive a scenario are constructed using the same tool set used to build the objects and behaviors of the experiment subjects.

4. <u>Executive Process-Interactive Control (EPIC);</u>

EPIC enables procedural cognition, motor control, and perceptual-motor interactions to be treated explicitly and parsimoniously in conjunction with formal hypotheses about supervisory executive cognitive processes and task-scheduling strategies. Precise computational models can be constructed to explain and predict reaction times (RTs), response accuracy, and other measurable aspects of people's overt behavior across various domains where multiple tasks must be performed concurrently.

5. Goals, Operators, Methods, and Selection Rules Language/GOMS Language Evaluation and Analysis (GOMSL/GLEAN3);

GOMS is a family of techniques proposed for modeling and describing human task performance. Most GOMS techniques are, at least partially, based on a simple cognitive architecture known as the Model Human Processor (MHP). This representation of human cognition consists of separate components for cognitive, motor, and perceptual processors (and associated buffers), as well as for long and short-term memory.

6. Human operator simulator (HOS);

Micro Saint simulation engine (see 8. Micro Saint) with the Human Operator Simulator extensions. The HOS extensions provide a mechanism to define a workspace associated task network. Built-in micro models of human behavior can be related to the operator and the control and used to dynamically modify the time to perform a task based on operator position. Micro models are functions that represent basic human actions.

7. Man machine integrated design and analysis system (MIDAS);

MIDAS contains tools to describe the operating environment, equipment, and mission of manned systems, and uses embedded models of human performance (e.g., vision, memory, decision making, and anthropometrics).

8. Micro Saint;

Micro Saint is a general purpose, discrete-event simulation software tool that is well established. It was developed under DOD funding and has been successfully used in numerous contexts including DOD, small businesses and Fortune 500 companies, and in many areas including human factors, ergonomics, health care, manufacturing, and the service industry.

9. Neural networks;

These general-purpose tools have been used to model many kinds of systems from financial analysis to oil exploration to brain networks to human performance. Some of the tools we examined for creating neural networks include:

a. NeuroLab - allows users to construct their own flexible and complicated artificial neural networks by simply clicking, dragging and connecting iconic blocks. The parameters of networks such as back propagation methods, learning rates, initial weights, and biases can be changed interactively in the dialog box of each functional block.

b. PDP++ - A neural-network simulation system written in C.

c. NeuroShell 2 - General purpose neural network construction and analysis tools. Includes 16 classic neural network architectures and the capability to design new architectures.

10. Operator Function Model expert (OMFspert);

OFMspert is the software implementing the operator function model. The operator function model (OFM) has evolved over a period of fifteen years and has matured to provide an increasingly robust design tool. OFM and OFMspert have been used to address a range of important but difficult human-system interaction issues in a variety of domains and for a range of applications. Domains include electronic manufacturing, aerospace, aviation, and medical systems. Applications include specification of human-centered operator workstations, intelligent associates, tutors, and, operations automation.

11. State, Operator, and Result (SOAR);

Soar is a general cognitive architecture for developing systems that exhibit intelligent behavior. Soar provides the fixed computational structures in which knowledge can be encoded and used to produce action in pursuit of goals. It has embedded in it a specific theory of the appropriate primitives underlying symbolic reasoning, learning, planning, and other capabilities that are hypothesize to be necessary for intelligent behavior.

12. Situation awareness model for pilot-in-the-loop evaluation (SAMPLE)

SAMPLE (Situation Awareness Model for Pilot-in-the-Loop Evaluation) was developed for evaluating subsystems and tactics in enhancing pilot SA during the course of air missions. The system consists of 1) a SA-centered pilot model; 2) a SA/performance metric generator; and 3) an interactive Graphical User Interface (GUI). The pilot SA model is implemented using Belief Network (BN) technology.

Preliminary Recommended Model

Of all the models surveyed, COGNET and SOAR initially appeared most promising. As noted above, all other models had specific strengths, but none appeared to provide the combination of generality, functionality, and technical support of these 2 models. Both COGNET and SOAR are rule-based architectures. Of the 2 models, COGNET has gotten wider use. It is a commercial product with an easy-to-use GUI and very good technical support. SOAR is still a non-commercial product with a more difficult interface, but it is well supported by an enthusiastic user/developer community, especially at the University of Michigan.

However, after considerable reflection we became convinced that, while existing models provide a rich source of suggestions, none were completely appropriate for the purposes of this effort. Many include esoteric details designed to incorporate basic distinctions in modes of processing, stylistic differences, or relatively small effects that would have very little impact on the overall cognitive processes of interest in the acceleration environment.

In view of this, NTI decided to synthesize a model specifically directed to the skills required in the acceleration environment. A start toward identifying these critical skills was made during a workshop held at Wright-Patterson AFB in 1995, and described by O'Donnell, Cardenas, Eddy, and Shaw (1995). The "critical" operational tasks that pilots considered are potentially vulnerable to degradation from G-forces are shown in Table 1.

TABLE 1

CRITICAL PILOT TASKS SUGGESTED IN THE AL/CFBS WORKSHOP

- Observe and kill bandit
    - Maintain sight
    - Maintain advantage
        - Manage energy
        - Achieve shot parameters
- Verbal
    - Communicate
- Motor skills
    - Initial missile avoidance
    - Detect second missile
- Visual
    - Acquire target
    - Recognize threat
    - Evaluate threat
        - Radar lock
        - Missile parameters
        - Bandit range
        - Etc.
- Awareness
    - Analyze situation
    - Check gas, airspeed, floor, etc.
- Cognitive
    - Develop plan (engage or leave)
    - Analyze bandit's maneuver

Of course, not all of these tasks are equally vulnerable with respect to G effects. For instance, the instant ability to recognize a threat and respond appropriately might be considered more critical after G-exposure than the skills required to monitor fuel. Therefore, it was necessary to consider which ones were most likely to be impacted by acceleration forces. This analysis was carried out by subject-matter-experts within NTI, and helped to focus our attention on <u>the most critical skills</u> in the fighter environment likely to be impacted by acceleration forces. These then formed the basis of both the cognitive model to be developed, and on the performance tests that would assess the nodes in the model.

Using this 'operationally relevant' data as a limiting criterion, our intent in developing a synthesized model of performance was to utilize as much information from existing cognitive models as possible. We were especially interested in model elements where there seemed to be an overlap or agreement among various models that a particular function or element was critical to cognitive/motor performance. In all cases, however, we were guided by the types of critical skills that are typically required of fighter pilots who might undergo extreme acceleration forces. In other words, the model should address the skills required to carry out the tasks identified in Table 1.

The starting framework of the model that we selected is shown in modified form in figure 1. This is a classical approach to cognitive modeling dating back many years (e.g., Broadbent, 1958). It simply views performance as a function of input (sensory/ perceptual) factors, central processing factors, and response factors. Of course, since none of these are elaborated upon, this general framework simply establishes what needs to be filled in to instantiate it. In one way or another, nearly all of the models surveyed above explicitly or implicitly adopt this approach. Our task, in Phase I, was to elaborate on each of the generic elements of this approach in order to produce a model that synthesized, or captured, essential elements of the more complex models, but was limited to the critical tasks of interest. The results of these efforts are summarized below.

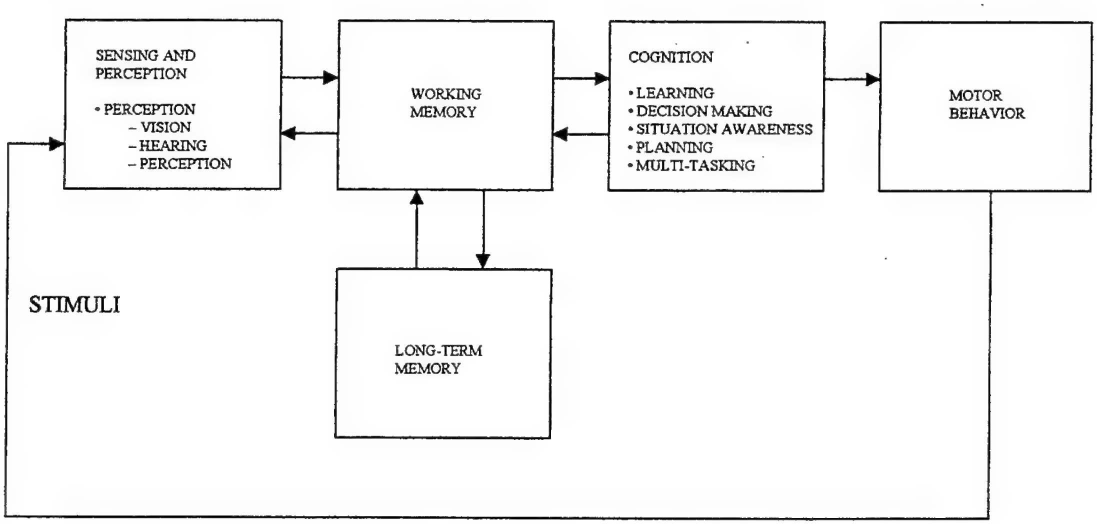

Figure 1. Generic model

Sensory/Perceptual Element

With respect to the sensory/perceptual element, it was decided for this initial effort to focus on the visual, kinesthetic, and vestibular senses. Auditory, tactile, and touch senses can be important in flight, of course. However, they are not typically considered to be uniquely degraded by G forces. Vision and vestibular inputs are critical in this environment. In view of this, the first element of the general model above can be subdivided into three critical sensory elements, as shown in figure 2. However, the perceptual aspect of this input must also be considered. In this context, the perceptual aspect represents a secondary interpretation or computation that is carried out on the raw sensory input. As such, it is not strictly independent of later processes, and forms a unique feedback/feed forward circuit dependent on sensory input and later processing. This is also shown in figure 2.

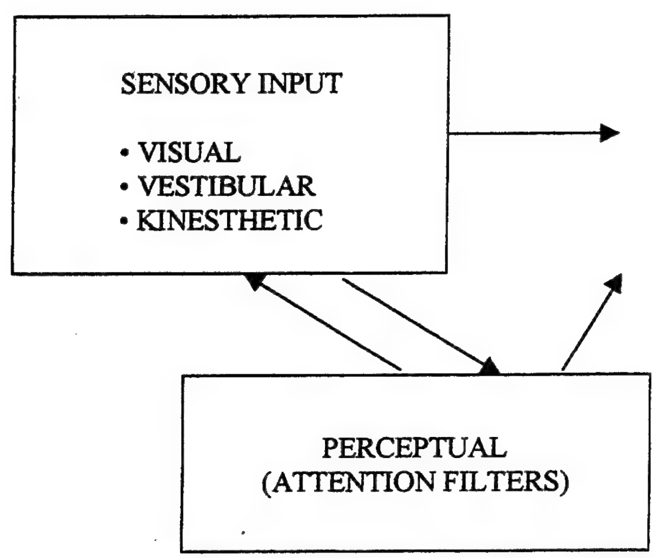

Figure 2. Sensory/Perceptual elements of the synthesized model

Central Processing Element

The "central processor" element in the general model above has received a considerable amount of attention in the cognitive literature, and has generated as much controversy as clarity. Early models spoke of "short-term memory" and often postulated elements within short-term memory such as buffers and rehearsal loops. Indeed, this appeared to be a useful approach, since factors affecting longer-term memory, such as interference and decay, could be explained. Later, the concept of "working memory" gained popularity. Working memory appears in many models to be synonymous with the contents of conscious processing. This concept allowed theorists to incorporate such functions as declarative versus procedural memory, and crystallized versus fluid memory.

Both terms are still currently used, although there is no universal agreement on their meaning or their operating principles.

For the present purposes, we decided again to take a very pragmatic view of these central processor functions. Our goal was to define those central processes critical to handling and interpreting the kind of sensory/perceptual inputs required from the pilot, and to do so in a way that could be easily modeled. In other words, fine distinctions between types of memory and modes of processing would be ignored, as long as doing so would not affect the kind and degree of performance prediction of interest to the acceleration environment.

In carrying this out, we loosely utilized a computer metaphor. We viewed **short-term memory** as analogous to the data that is put into the system. Thus, the sensory/perceptual outputs constitute the raw material of short-term memory. There may or may not be interference or decay functions that occur at this stage, but the possibility of incorporating these into the short-term memory element should certainly be maintained. Parenthetically, it should be noted that the ability to test the concept of short-term memory when viewed in this way is problematical. Any test that requires the subject to make a response, by necessity, is going to involve more complex levels of processing, and therefore will be somewhat confounded with such levels. However, we believe it will be possible to develop probes that will, for all practical purposes, indicate the functionality of this short-term memory module.

We view the concept of **working memory** as being analogous to the program that is written to process the elements in short-term memory. In other words, this element is the central processor. In this, we tend to agree with Kyllonen and Christal (1990) that working memory capacity may be considered equivalent to what is usually thought of as "reasoning ability". On the one hand, it can do nothing without appropriate input from the sensory/perceptual, and/or short-term memory systems. (We ignore, for the moment, input from long-term memory – as well as the implications of "implicit memory" experiments). On the other hand, no real information processing in the traditional sense can be carried out without this system. In the most general sense, we hypothesize that this working memory module is relatively undifferentiated at birth, although there is almost certainly neural pre-wiring disposing it in specific ways. The "programming" of the working memory system is a function primarily of learning (procedural and declarative memory), and secondarily of personality pre-dispositions, and the neural wiring mentioned. Such concepts as controlled and automatic processes have been used to describe various levels of attention requirements in this system, although for the present purposes this distinction probably is not crucial since, in the pilot, most required actions will have achieved some level of automaticity.

We view this system as critically important to the performance predictions that will be made in the G-TOP system. Again, there is considerable controversy concerning the exact composition and operating principles of this working memory. Many of these are concerned with details of how the system functions (e.g., whether through a "blackboard"

structure or through direct module-to-module communication). These questions are not of primary concern here. Rather, the concern is in assuring that the "output" of the model matches real-world behavior. Thus, we focused on models that asked questions concerning the critical sub-elements of working memory. These involve such questions as "Are there separate 'knowledge modules' that feed into working memory, and if so, what are they?" and "How many distinct working memories are there?" (Shallice, 1988; Baars, 1988).

Essentially, we adopted the view that there are such "knowledge modules" that essentially consist of learned procedures that specifically operate on a class of inputs from short-term and long-term memory. These can loosely be considered "skills" that constitute the combination of short- and long-term memory content and working memory process. Of course, these modules do not operate in isolation. Attentional processes, motivation, emotions, and many more factors influence them. However, most of these can be assumed to be optimal in the average pilot. Most critically for the present purpose, the knowledge modules will be influenced by the physiological condition of the pilot during and after G exposure. Therefore, we can focus on these modules or skills in relative isolation from the large number of confounding factors that complicate more extensive models of performance.

Many separate knowledge modules probably exist. Prominently mentioned candidates include syntactic and spatial analysis, language, mathematical ability, and visual-motor control (Baars, 1988), as well as logical reasoning, situation assessment, decision selection, and action monitoring. Again, most of these are not of primary interest in the acceleration environment since they are probably not typically involved in the "critical tasks" required from the pilot during or after G exposure. Mathematical facility, logical reasoning, verbal fluency, and syntactic analysis are examples of skills that may be important in other contexts, but not in the present one. On the other hand, inspection of the tasks that are critical to the pilot during or after G exposure indicates that such elements as visual-motor control, spatial relations, situation assessment, decision selection, and action monitoring are critical. Therefore, we include in the model working memory sub-elements designed to capture these functions.

The role of long-term memory processes in interacting with working memory has already been mentioned. We envision long-term memory as being integrally involved in working memory, since it supplies the schemas, recognition capacity, and experience base that permits such functions as response selection to operate efficiently. Therefore, in the model for this effort, we indicate a significant role for long-term memory. However, we also conclude that, although one could perhaps test long-term memory in the abstract (e.g., simple tests of retrieval, etc.), for the present purposes it is more important to test the dynamic aspects of long-term memory. That is, we desire to test how efficiently long-term memory functions in <u>coordination</u> with working memory. Therefore, in the section below describing the development of the revised test battery, long-term memory functions are probed through more active tests involving rapid decisions and response selection that appear under working memory itself.

The preliminary model developed during Phase I is shown in figure 3. We label this model "preliminary" because it represents our first approximation to a final model. We believe that, with further evolutionary development, this model will provide the underlying structure that will be able to accept input from the test battery described below, and will also provide output appropriate to further analyses.

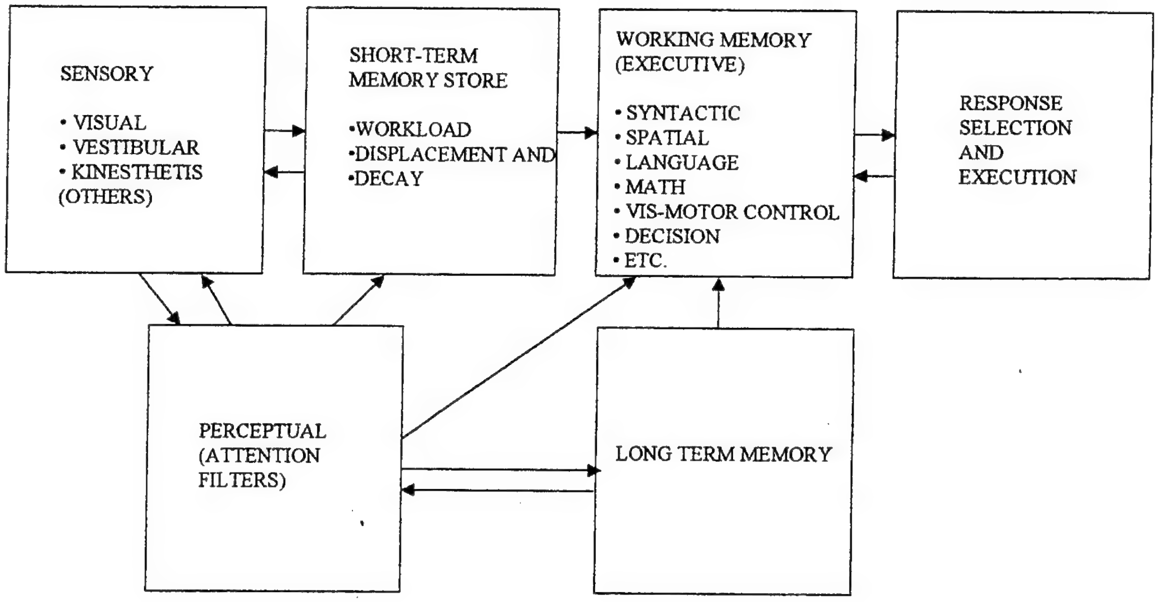

Figure 3. Human Performance Model Developed in Phase I

DESIGN OF THE G-PERFORMANCE ASSESSMENT SIMULATION SYSTEM (G-PASS)

Introduction

One of the goals of this Phase I was to design a test battery that would be capable of supplying performance data as input to the above model. The starting point for that effort was a previous development carried out by NTI, Inc. In 1995, we designed a battery of performance tests for the acceleration environment as part of an Air Force Phase I SBIR. This battery was called the Acceleration-Performance Assessment Simulation System (A-PASS), and was described in some detail in the final report for that project (O'Donnell et al, 1995). However, since the Phase II for that program was not funded, the test battery was never implemented or tested. In the current Phase I effort, the original A-PASS test recommendations were reviewed in view of the human performance model developed and explained above. Some tests in the original battery were found to be inconsistent with the new battery, while some others simply required slight modifications to be useful. The resulting modified battery is called the **G-Performance Assessment Simulation System (G-PASS)**, and is described below.

As in the earlier Phase I SBIR carried out in 1995, the current Phase I defined a battery of tests designed to be appropriate for the centrifuge environment. However, the approach in the present case was somewhat different. In the development of the original A-PASS battery, we were guided by an attempt to formulate tests to address critical flight tasks. Experienced pilots, including active Air Force fighter pilots, established the basic requirements that resulted in selection of the particular test procedures in the A-PASS battery. In the earlier effort, there was no attempt to structure the battery around a theoretical model of human performance. In the present Phase I, the underlying model provided a coherent framework on which to select tests, and introduced some new elements into the desired test characteristics. Therefore, we approached the task as if it could result in an entirely different set of test than those selected for the A-PASS battery.

In point of fact, however, it soon became apparent that many of the originally proposed tests also fit into the human performance model. In retrospect, this is not surprising. The basic skills sampled in the A-PASS battery were chosen precisely <u>because</u> they represented elementary dimensions of human performance. The final test battery conceptualized in the present Phase I, therefore, contains eight of the test procedures from the A-PASS battery, although some of these have been modified in some cases to better fit its intended use in the performance model. In addition, four new test procedures have been added to the new battery and constitute the G-Performance Assessment Simulation System (G-PASS).

TESTS BASED ON THE A-PASS TEST BATTERY

Test 1. Perception of Relative Motion

General description of the test

One of the more important skills required of the pilot is to be generally aware of the relative positions of one or more other aircraft with respect to his or her "own ship". It is therefore desirable to probe the pilot's ability to demonstrate these skills. With respect to the model of human performance, this skill primarily involves the visual sensory/perceptual system. It was decided to include a "formation join-up" task of the original battery in the new G-PASS system, essentially without modification.

Originally, we considered using a high-fidelity air-to-air engagement model to probe these skills. However, actually implementing and scoring such a model could become extremely complex. Therefore, we searched for a task that would make the same skill demands on the pilot, while constraining his or her options. It was suggested by our pilot consultants that the operational task of joining up with a formation or another ship requires the same types of manual skills as any other air-to-air engagement. The difference is that the goal is clearly defined or definable (i.e., joining the formation as quickly and efficiently as possible) and the maneuvering options can likewise be mathematically described. In addition, the maneuvering of the other ship can be pre-described or determined, thus eliminating the need for artificial intelligence to be built into the system.

For these reasons, a formation join-up task was chosen for inclusion in the original A-PASS test battery, and the same reasons appear valid for the new battery. Generally, the task described below and shown in figure 4 involves having the subject "fly" a target to another moving target. The subject will use throttle and stick inputs to control one of the targets in three dimensions (up-down, right-left, or a speed dimension). The subject's task will be to "join up" with the moving target as quickly as possible.

The major parameters that will be manipulated in this task are the initial starting positions of the two targets, and the maneuvering of the moving target. The subject will first see both targets on the screen in one of eight pre-defined starting positions. The goal will always be to make the two targets just touch each other (i.e., the goal is to join the two targets as quickly as possible without "crashing"). The moving target will describe one of eight trajectories, ranging from a straight line in the horizontal dimension to a rapid "jinking" maneuver. Timing will commence with the onset of the stimulus materials, and will end when the two targets touch each other. If the targets approach each other too rapidly, they will "crash", and this will also serve as a data point.

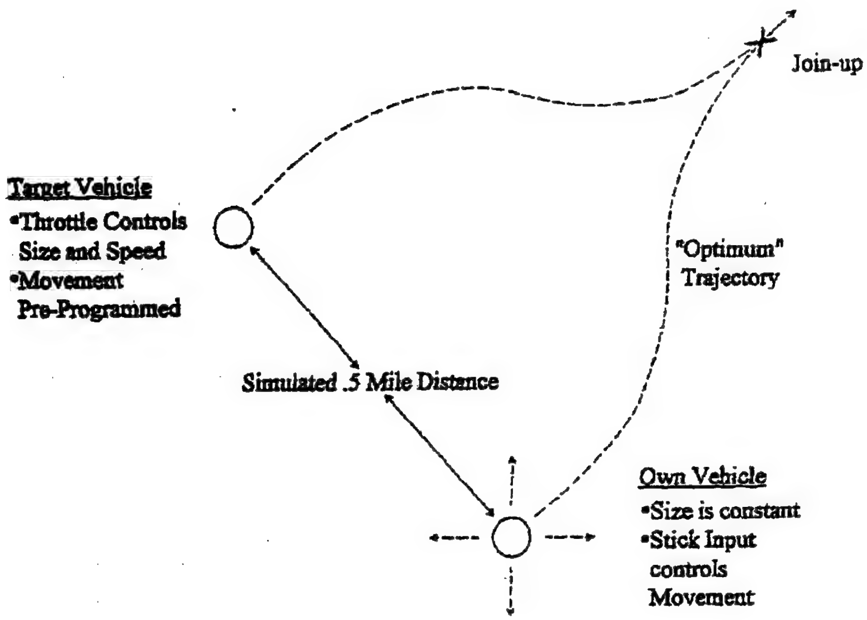

Figure 4. The Perception of Relative Motion Task

Detailed test specifications

Individual trials on the task should average between five and fifteen seconds. The number of trials required for stability will, of course, need to be determined after the test is actually constructed. However, we would expect that a stable baseline could be obtained in less than five minutes of training, encompassing ten to twelve trials. We also feel that after adequate training, as few as three trials might yield stable results in any experiment.

The stimulus elements on the screen should consist of two circles representing aircraft (see figure 4). One circle ("own vehicle") will remain stable in size (approximately .75 inches), and stationary. However, control movements by the subject will cause the "other ship" to appear to change its relative position, using dynamics which approximate the response characteristics of an F-16 aircraft. The initial distance between this icon and the second one should simulate approximately .5 mile, or whatever distance would permit a potential link up in approximately five to ten seconds, assuming a terminal closure speed of about 10 kts. at 1000 feet.

The second circle will also represent an aircraft ("other ship", or target). This target should be preprogrammed to move on its own. This movement should generally be relatively slow, to approximate the expected visual motion of a friendly ship that is waiting for a formation link up. In most cases, the trajectory of this motion should either

be a straight line or a fairly gentle curve. Periodically, however, this aircraft may make a sudden move up or down, with a change in apparent speed. This situation should model the situation in which the aircraft was required to make an evasive maneuver, or simply committed an error. In addition to this pre-programmed movement, the target circle should appear to move in response to the subject's own control inputs. In other words, the simulation should appear as if the pilot were observing the target circle from inside his or her own cockpit.

The size of this second icon should be controlled by the subject's throttle. Thus, the interaction between the joystick and throttle will determine an apparent "closure rate" between the two icons. As the subject flies toward the target aircraft, it will become larger (and appear to move faster) in proportion to how close the two objects are. This throttle control, therefore, obviously simulates the speed of one's own ship in approaching the other target.

The subject's task is simply to "fly" his or her own ship to the point where the "aircraft" is a predetermined distance from the target "aircraft". This is to be done as rapidly as possible without, of course, crashing into the other aircraft. The task will therefore require the subject to rapidly decide on an optimal trajectory that will place the aircraft together. Once this determination is made and an initial flight path is initiated, the subject must determine the appropriate speed at each point along the flight path. Vigilance must be maintained, since the subject will be aware that the target aircraft occasionally makes unpredictable moves. If such a move should occur, the subject would need to recalculate a new trajectory and speed to minimize the link up time. (It is, of course, possible to introduce additional requirements or decision tasks. For instance, if the target aircraft makes a sudden jinking maneuver, the subject might be required to abandon the link up, and make a similar jinking maneuver.)

The precise definitions of an "optimum trajectory" for various types of join-up will be determined during actual construction of the test, in consultation with pilot consultants and Government personnel. Decisions concerning these involve a number of factors that can only be determined after some prototype demonstration of the test is available.

The principal scoring dimensions of this task will be the time taken to link up, and the accuracy of the link up. We do not anticipate calculating such things as RMS error from the "optimal" trajectory. Nor do we anticipate collecting measures of such things as simulated fuel consumption. These, however, could be considered as potential metrics from this task.

<u>Test 2. Precision Timing Task</u>

<u>General description of the test</u>

The essential skill demand in this category is that the pilot visually monitors a changing situation, and decides at some critically identified point to initiate a motor action. The

piloting demands which require this skill involve precision timing (general timing ability), especially visually directed precision timing. These include such things as decisions on weapons release in both air-to-air and air-to-ground situations, flare decisions, formation flight, and decisions to abort landings or other activities. The essential skill demand is that the individual visually monitor a changing situation, and decide at some critically identified point to initiate a motor action. With respect to the performance model, this task appears to require visual/motor integration at the level of working memory, but also has elements that extend over all of the nodes in the model to varying extents.

To probe this type of skill, a task was developed in which the individual must monitor what appears to be a rapidly moving target for a brief period of time (figure 6). At some point in the path of the target's motion, an indicator appears along the target's path. The subject's task is to press a button on the stick when the indicator reaches that point. The distance and/or time error of the subject in precisely stopping the target at the right time will constitute the basic measurement parameter of this task.

<u>Detailed test specifications</u>

The subject will see a curvilinear pattern of dots that will appear to move across the display screen. Figure 5 illustrates this initial screen. It will be noted that somewhere in the pattern, there is an indicator (here shown by a line labeled "Sample Target Point") designating some specific point in the path of the dots. Shortly after the appearance of the initial screen, the dots will begin to light up in a way that will appear to the subject to be a moving light through the curvilinear pattern. This light will move rapidly, although at different speeds on each trial. The subject's task will be to monitor the progress of the light toward the indicator, and to stop it precisely at the indicator by pressing one of the buttons on the control stick. Once a trial has been completed, the subject will be given a brief period of time to inspect the results.

The initial screen presented to the subject should contain at least 50 dots arranged in a semi-circle, with its end points located one inch from the left and right side of the screen, and one inch above the bottom of the screen. The top of the circle should reach to approximately two inches below the top of the screen. The dots should be equally spaced, and no more than 1/32 inch in size.

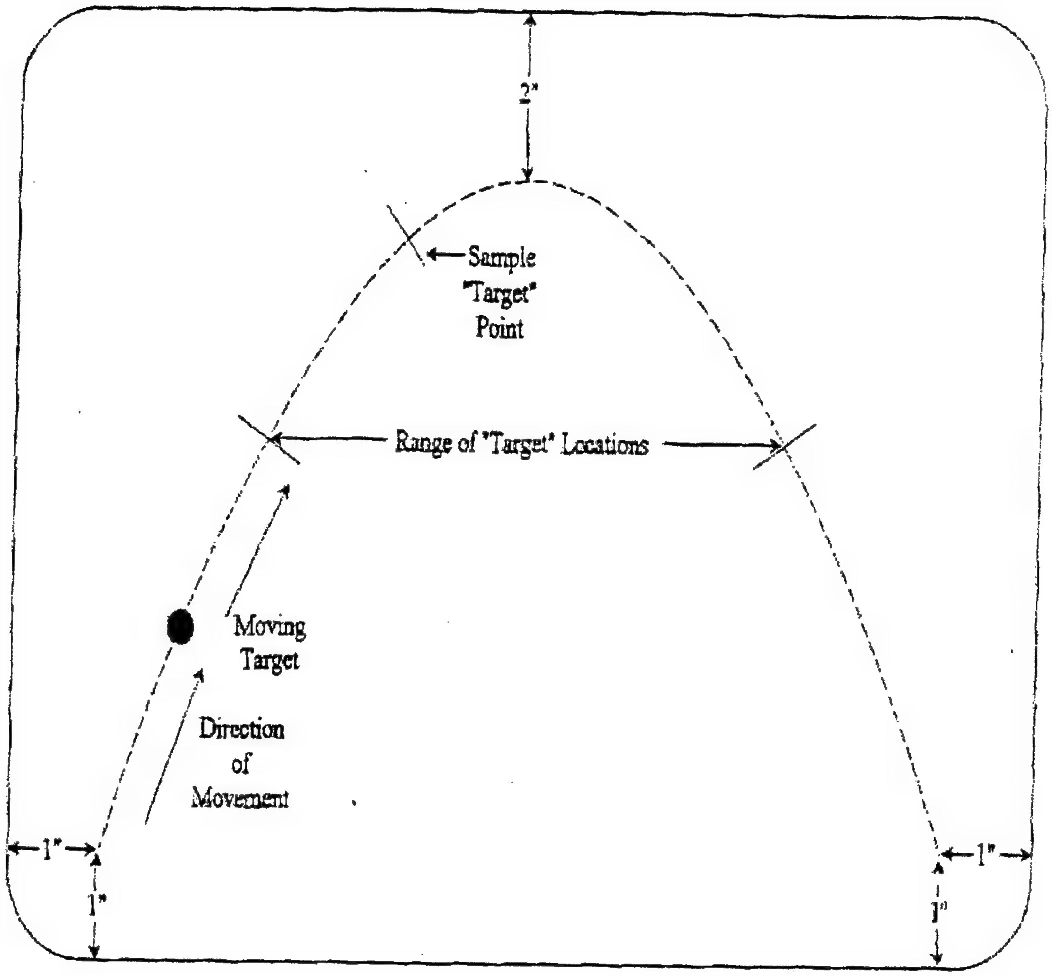

Figure 5. The precision timing task

The subject will see what appears to be a moving light which will go from left to right around the path described by the semicircle. This apparent motion will be achieved by having each dot position "flash" in a rapidly moving sequence. In other words, starting from the extreme left hand side of the semicircle, each dot will appear to "light up" or brighten in sequence until the sequence ends at the extreme right hand dot of the semicircle.

The speed at which the light moves around the semicircle will constitute one of the major independent variables in the test. The minimum speed, or the maximum inter-flash interval, should be calculated based on the intensity of the screen ultimately used. This minimum speed will be the slowest interval that will produce the phi phenomenon. It would be expected that this would be in the range of 750 milliseconds. This would yield a trial with a maximum length of approximately 2.5 seconds. The fastest speed at which the apparent light will move will also be determined empirically after the test is constructed. However, it will be defined here as the fastest achievable speed at which

subjects can still give reliable responses. It would be estimated that this fastest achievable speed will be in the range of 40 to 80 milliseconds.

To achieve the capability of having the speed of the moving light change, the basic program directing the interflash intervals should permit specification of an acceleration or deceleration factor. The default option will be the constant speed. Again, it will be necessary to determine experimentally what types of acceleration changes subjects can respond to reliably. However, it would not be expected that, given the difficulty of the task, a wide range of such acceleration variation will be able to be used. For purposes of this test, it will be sufficient to have a few types of acceleration variations that are easily recognized by the subject under nominal conditions.

The indicator that will tell the subject where the light should be stopped may be located anywhere within the center 1/2 of the semicircle (indicated in figure 6 by the "Range of Target Locations"). Eventually, it would be expected that a standardized protocol will be developed specifying exactly what positions should be utilized. In its most generic form, however, this parameter will be left to the experimenter's discretion. In other words, the experimenter will be able to designate exactly where he or she would like the indicator to be placed. The indicator itself will consist of a 1/2 inch line which will fall between two adjacent dots.

A testing sequence should consist of a minimum 30 trials, based on the need for reliability in this type of measure. As always, more trials are desirable, and there will be no limit on the number of trials an individual experimenter might utilize. An individual trial might last between .75 and 2.5 seconds. If the test is administered continuously (i.e., with no long breaks) the intertrial interval should be constant, and should be at least 1.5 seconds long. Thus, a minimal test (30 trials) would be expected to last approximately 1.5 minutes.

Two basic types of scores will be obtained in this test. The simplest score will consist of the absolute distance error between the indicator and where the subject actually stopped the lights. Obviously, the direction of the error (short or long) will also be noted. The summary statistics for this measure will be the mean arithmetic error, the mean absolute error, the standard deviation of both of these, and the RMS error. These will be calculated over all trials.

The second type of measure is based on the fact that there will be different levels of difficulty in many of the trials. Slower moving lights should reasonably be easier to project accurately. Although this feature of the test adds some degree of flexibility to the experimenter's analysis options (e.g., it may be found that certain stresses affect only high speed light perception and timing), it makes interpretation of the simple statistics described above somewhat problematical. To alleviate the situation, a statistic based on time will also be calculated. In this, the subject's distance error on any given trial will be converted to the amount of time by which he or she was in error. In other words, if the subject was two centimeters short of the target, and if the light was "moving" at the rate

of ten centimeters per second, the subject should have waited .2 seconds longer. This case would yield a score of .2 seconds. In another case, if the subject was again two centimeters from the target, but the light was "moving" at the rate of five centimeters per second (i.e., slower than the first case) the subjects error time was .4 seconds. This scoring system therefore accounts for the fact that the second task was presumably easier than the first. Therefore the same distance error should result in a poorer score in the second case than in the first. This calculation will be carried out automatically, and the same statistics calculated above will be calculated for this type of data.

Test 3. Motion Inference

General description of the test

In the combat situation, there are instances where the individual must perceive and register the motion of an enemy or another object, then must turn away from direct visual perception of the object briefly. However, the absent object's motion must still be processed in order to know where it should be when it is again attended to. In such cases, the individual must infer motion based on a previous perception of motion. This must frequently be done while other tasks are being performed. Clearly, this places a considerable burden on working memory, especially as it interacts with short-term memory. Response selection and timing are also critical in this case.

Although this is a situation that occurs with reasonable frequency in flight, it is not an easy one to quantify. Attempts to do so have essentially utilized a moving target that, at some designated point, either disappeared or was hidden from view by an obstacle. The subject had to estimate when the object, moving at a constant rate of speed, would reappear from behind the obstacle, or would "hit a target" (Broach, Goldbach, Weltin, et al, 1993). This approach probably constitutes a good measure of the basic skill. However, it allows the subject to focus completely on this one task, and subjects sometime develop "strategies" which confound the measure (e.g., they sometimes "count beats" or sing music in time with the motion). This confounds interpretation because it is not usually the way the task has to be done in the real world. In those situations, the individual is frequently preoccupied with at least one other task while making these inferences. Therefore, it was desirable to incorporate some form of distracting task during the period of time that the subject is making inferences.

The A-PASS test designed to address this type of performance requirement involved having the subject view a moving light traversing a curved path (figure 6A), and this task appears to be appropriate for the present battery. The light disappears at some point, and the subject's task is to determine when it would have reached the end of the curved path. A distracting task is introduced during the time estimation interval (recognizing vowels in a string of four letters), simulating distractions that typically occur in the real world (figure 6B).

Essentially, the subject will see a moving light traversing a curved path (as in Task 2 above). Approximately half way through the curve, the light will "go out". The subject's task is to determine when the light, moving at a constant speed, would have reached the end of the curved path. In other words, the subject must infer how long the light would take to reach the end of its path. The response required will be a button press when the subject believes the light would have reached the path.

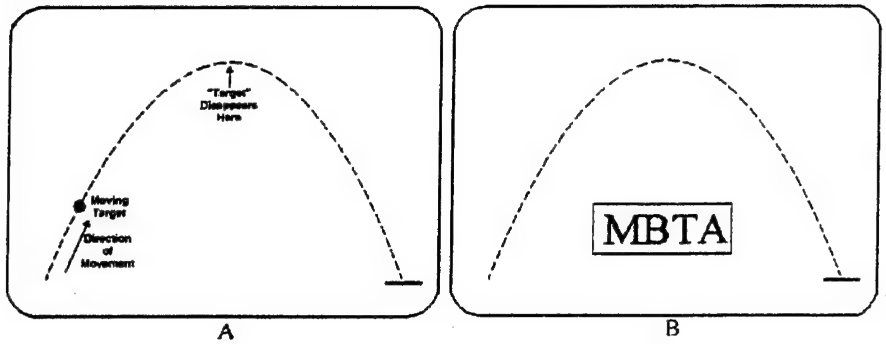

Figure 6. The motion inference test: A – the basic motion screen: B – the motion screen with the distracting task

The distracting task will be a simple "semantic" one. When the light goes out, a series of four letters of the alphabet will appear on the screen. The subject must immediately decide whether any of the letters are vowels. In other words, this interpolated task will act as a distracter to the subject in estimating the inferred motion. In this way, the subject will be precluded from using methods such as counting, tapping, or singing to infer the motion. It also simulates more closely the tasks required in the real world.

<u>Detailed test specifications</u>

This task will utilize the same basic stimulus environment as the precision timing task described above. In other words, the path traversed by the "light" will be the same as in that task. The subject will initially see the dots describing the path that the light will take. The "light" will consist of a brightening and enlargement of sequential dots at a rate sufficient to generate the phi phenomenon. The light will appear to move at a constant speed for each trial. However, the speed will differ from trial to trial. Again, the range of speeds at which the light will appear to travel will have to be determined experimentally. Nominally, speeds ranging from .5 to 2.5 inches per second will be used in initial trials.

As soon as the apparent light reaches the center of the semicircle, it will appear to go out. Immediately, four alphabet characters will appear at the center of the arc described by the semicircle, as shown in figure 6B. These will subtend approximately .5 minutes of visual angle, in order to be large enough not to be affected by small visual acuity changes. They will be located not more than .5 inches from each other in a straight line. The subject's task will be to indicate, as quickly as possible, whether the series contains a vowel or not.

This will be done with the left hand, using a designated button on the throttle. Once this is done, the subject will then estimate when the light would have reached the extreme right-hand end of the semi-circle, and will indicate this by pressing a "fire" button on the stick.

The selection and distribution of letters for this Subtask is, of course, critical. Different letters should be used on each trial. However, certain parameters will be established to ensure that, within each series of 32 trials, the following conditions will occur:

1) There will be a 50-50 split in the probability of a vowel appearing in the sequence.

2) Only one vowel per series will be present.

3) The vowel will appear in each position an equal number of times.

4) The same vowel will not appear in two contiguous trials more than twice.

5) No "run" of more than four trials consecutively will be permitted for either vowel or non-vowel conditions. Beyond this requirement, no randomization formula will be used.

To achieve the above control over the parameters of the presentation of the task, a "canned" set of 32-trial sessions should be produced. At least 20 such sessions should be developed and available for presentation to the subject.

It is estimated that a single trial for this task will last between two and four seconds. NTI recommends at least a three second intertrial interval to permit subject feedback and rest. Thus, each trial will occupy between five and seven seconds. Nominally, at least 32 trials should be given in a test session in order to minimize variability. Therefore, it would be estimated that the test overall might last, on the average between 3 and 5 minutes.

The type of data produced by this task is very similar to that produced by the precision timing task. Error scores will be available in the form of either distance or time between the actual target point and at the point at which the subject responded. Therefore, the data will be treated in the same way as described for the precision timing task.

Test 4. The Pitch/Roll Capture Task

General description of the test

A critical survival skill in air-to-air combat involves the ability to rapidly position the aircraft in order to move a bandit into a specific location relative to your aircraft's aiming devices. Essentially, the pilot must recognize the bandit's relative position, and then must make a rapid correction in his or her own position and attitude in order to gain the

proper aiming advantage. This rapid maneuver might be in the vertical relative to the aircraft ("pitch capture"), or laterally ("roll capture"). These terms refer to the required control input, pitch or roll, that will bring the bandit into the desired position. Delays or errors in doing so will obviously have an impact on the outcome of the air engagement. On the other hand, if the pilot demonstrates a normal ability to carry out this type of task, it should be possible to assume that the person will be able to carry out other, less rapid "capture" tasks (e.g., formation flight).

This task probably involves overlearned long-term memory functions to a great extent, but also may involve spatial working memory, response selection, visual/vestibular interactions, and visual-motor control. To probe the skills required for this type of task, the A-PASS battery contained a medium-fidelity simulation of the actual pitch and/or capture as it would be carried out in an F-16 aircraft (see figure 7). The subject will see a crude front cockpit simulation, and will be required to be performing a routine, easy cockpit task (e.g., flying straight and level, or adjusting a radio frequency). At some random time during the trial, a target will appear in some location around the cockpit field of view. The subject's task will be simply to move the target as rapidly as possible into an instructed firing position, using the control stick. Scoring will be based on the time and accuracy of the movement.

Detailed test specifications

The initial screen viewed by the subject will consist of a forward cockpit view as shown in figure 7. This will present sufficient information for the subject to "fly" a straight and level path. Subjects will do this for a random amount of time, up to one minute in duration. At some point during this time, a "target" will appear on the HUD. The target will appear either directly above the aiming reticule, or offset to the side. In either case, the subject's task is to make a control movement of the stick, as quickly as possible, which brings the target into the aiming reticule. In the case where the target is directly above the reticule, this would involve pulling the stick back (pitch capture) to simulate raising the nose of the aircraft. If the target appeared directly to the side, the required response would require first a roll maneuver, and then a pitch maneuver. In all cases, the subject will be required to press a "fire" button when he or she has achieved "capture" of the target.

At least 32 "standard" initial positions of the target will be pre-programmed for this test. The level of difficulty on a given trial will differ depending on whether single or dual movements are required to effect a capture. The researcher will therefore be able to customize the level of difficulty of the tasks to suit individual research requirements, and to assure that the difficulty level of tests done at various times are equated.

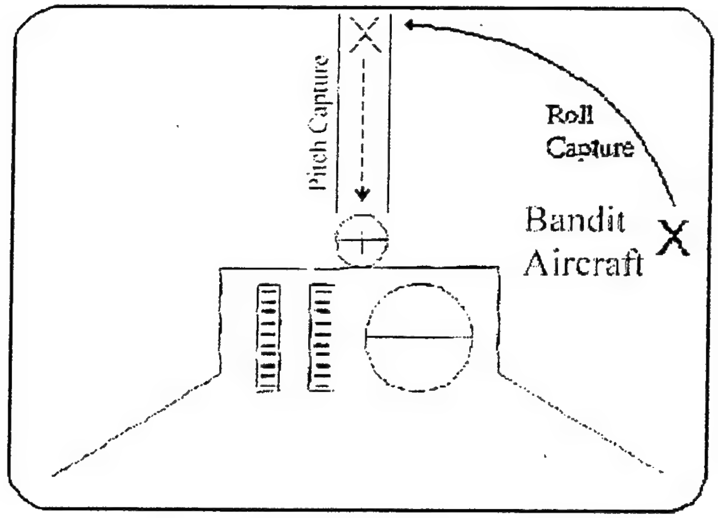

Figure 7. The pitch/roll capture task

A single trial in this test would be expected to last no more than one minute, with an average of slightly more than 30 seconds. Since this is a reasonably simple skill that is being tested, it would be expected that stability would be reached with one to two hours of training. Reliable results could then be expected with as few as 15 trials. Therefore, the duration of the test would be expected to be between 10 and 20 minutes. This should permit calculation of performance on this skill under a wide range of difficulty levels.

Scoring of this task will be straightforward. The basic dependent variable is the time taken to "fire" - in other words, the time taken to perform the capture. Of course, it will be necessary to modulate this measure by some estimate of the accuracy of the subject when the "fire" button was pressed. This will be done in a fairly crude way in the sense that, if the target was within the reticule when the fire response was made, the time will be accepted. If the target was not within the reticule when fired upon, the task will be considered "failed" because the target was missed, and the subject will be considered "killed".

Test 5. Gunsight Tracking Task

General description of the test

A relatively high fidelity simulation of the gunsight tracking task carried out by the pilot was recommended by the consultant pilots in the original A-PASS battery. However, in the interim period, NTI has gained experience in implementing a flight-profile technique the closed-loop centrifuge at Brooks AFB (i.e., the pilot must follow a prescribed path in the sky). Therefore, we have decided to use this basic experience to modify the originally proposed task for the G-PASS battery. The display seen by an F-16 pilot during a gun fight will still be retrained, but the trajectory (path) of the "enemy" will be

modified to reflect the dynamics of the Wright-Patterson centrifuge, and the results we have seen at Brooks AFB. The gunsight will be controlled by the subject, using the stick of the simulator, thus simulating moving one's own aircraft to bring it into firing range. The control functions that drive the "target's" movements will be precisely determined, yielding a large number of precision metrics dealing with the pilot's ability to capture the target.

Detailed test specifications

Since this task will be customized to the display characteristics of the Wright-Patterson centrifuge, exact test specifications cannot be determined without further input concerning anticipated modifications to that device. Essentially, however, the stimulus presentation to the subject for this test will consist of an out-the-window view from a fighter aircraft, and the interface between the display and control inputs will be supplied by a relatively high-fidelity aeromodel of the F-16 aircraft. In other words, the subject will control the aircraft in the same way it would be done in an F-16 aircraft.

The stimulus display for a test will include a "bandit" target aircraft that follows a pre-determined path. This aircraft will always remain at a fixed distance from the subject's aircraft (i.e., if the subject increases speed, the bandit will appear to also increase speed). The subject's task will be to maneuver to bring the moving "target" into the gunsight pipper, and then to "fire" the weapon for the entire time the target is within the pipper.

Specifically, the target to be used for this test should be an aircraft icon that "maneuvers" in a realistic fashion. Within the constraints of realism, target motion should be programmable by the researcher to reflect the research question being addressed (e.g., the effect of certain G-maneuvers). However, a default option will be supplied in the form of a "canned" scenario of target motion. The forcing function for this canned scenario will be a sum of sines input. The experimenter will determine the duration of each tracking trial through programming of the scenario. The default scenario will be at least three minutes long. However, the program should allow the experimenter to select any shorter duration.

Since this is essentially a pursuit tracking task, basic scoring will be straightforward. RMS error, absolute error, and time on target measures will be provided immediately after each run. In addition, however, a series of measures will also be obtained for the subject's accuracy in "firing" the weapon. These will consist of the absolute firing time where the pipper was "on target", the number of such accurate firings, and the duration and frequency of inaccurate firings. It should be noted that in these forms, the data provide directly useable input to the operational commander independent of their applicability to a human model. Any decrement in weapons release point from this test could be directly translated into CEP or damage assessment.

Test 6. Peripheral Visual Monitoring

General description of the test

In view of the large literature base on the use of peripheral lights as an indicator of G-LOC, the A-PASS battery recommended two types of peripheral monitoring measures: detection of a peripheral signal, and interpretation of the signal. Fortunately, peripheral vision is a critical sensory/perceptual element in the human performance model. We therefore decided to retain the originally proposed peripheral vision test in the G-PASS battery. The basic apparatus for this testing is shown schematically in figure 9. It will consist of a semi-circle of lights consistent with the "peripheral lights" apparatus currently used in the centrifuge at Wright-Patterson AFB. If only the light stimuli are used, the subject's task will be to detect when a light "flashes". This, of course, is different from the way peripheral lights are normally used in the centrifuge. There the goal is to assess G-LOC. In the present case, the goal is simply to assess peripheral vigilance.

In addition to the light stimuli, however, there will be vertical bars containing arrays of "dials" located to either side of the subject. These vertical bars will be moveable within the subject's range of peripheral vision. The subject's task will be to monitor these dials in a variety of ways. Various conditions closely simulating real-world tasks can be envisioned. For instance, in the nominal conditions, all the dials would point in one direction. The subject would be required to detect when any dial deviated from this nominal condition. He or she will have to monitor the peripheral information while carrying out some function that requires foveal vision. In other words, this foveal task will insure that the pilot must acquire the peripheral information non-foveally. The foveal task will be one which is reasonably easy to perform, but which requires constant attention. The dials will be constructed in such a way that the experimenter will be able to re-configure them easily in order to simulate any type of data input to the subject. Data from this test will be used to infer certain aspects of situation awareness/attention allocation in the pilot.

Detailed test specifications

The exact dimensions of the apparatus to be constructed for this task will depend, of course, on the specifications and limitations imposed by Government personnel as a result of the centrifuge requirements. However some approximations can be made at the present time. It can be estimated that the semi-circular light bar will have a radius of roughly 3 feet. It would be anticipated that the bar itself would be approximately three inches high, and would have a "carrier bar" attached to its rear surface to provide a slide for vertical bars that will be the carriers for the peripheral lights and displays. These vertical bars should be approximately two feet high by six inches wide. Four rectangular holes will be spaced equally along the height of the bar, as illustrated in Figure 8. These holes will be approximately four inches by two and half inches, or whatever size would accept standard aircraft peripheral displays (or simulated models of those displays).

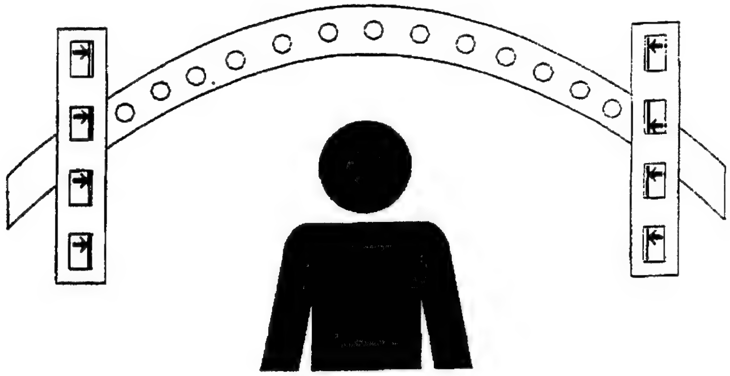

Figure 8. The peripheral monitoring test

The pre-drilled holes in the vertical bars will accept LED elements that should be able to be programmed to "flash" on a given schedule, or at a given event (e.g., a G-level). For the display dials, it is anticipated that at least two different types of dials will be used in the present task. One will consist of circular dials with pointer indicators. The other type of dial will consist of strip indicators with a needle capable of moving up and down the dial. Eight electro-mechanical indicators of each type will be fabricated in such a way that they can be secured through the pre-drilled holes. These indicators will be controllable by the computer program in such a way that each dial can be independently moved.

In the actual testing situation, it would be anticipated that the default conditions would involve two general types of stimulus presentation. In the first, all dials on both sides of the semi-circle will be placed in the same position. The subject's task will be to detect when any dial moved away from that position. In the second general condition, dials would be constantly moving, but within a "normal" range of movement. The subject's task will be to detect when any dial goes outside of this normal range (e.g., when any dial indicated a danger situation). Of course, several variations in these two basic approaches are possible, and will be programmable by the experimenter. The subject's response to an "out of bounds" condition will be to perform a control movement (e.g., a button press, or a stick movement) that will cancel the out of bounds condition.

The basic measurement parameters of this task will consist of the time between the occurrence of an out of bounds condition and the subject's response. It will also be possible to specify a "time-out" value that will suggest the subject simply did not see the anomaly. This would simulate catastrophic failure to attend to the peripheral information.

In order to assure stability in this type of task, it is anticipated that a short (e.g., 30 minute) training time will be required. In that amount of time, the task essentially will be over learned, since individuals have considerable experience in monitoring peripheral information of the type to be presented here. Once stability is achieved, it would be anticipated that no more than thirty well-balanced trials will be necessary to obtain a reliable performance from the subject. The duration of each trial could be as short as ten seconds. In that case, the average time for the appearance of an out of bounds condition would be five seconds, with a range from one to ten seconds. Of course, the experimenter should be free to adjust this time at will, and it would not be unrealistic to introduce individual trial times as long as one or two minutes.

The "central" foveal task to be presented to the subject can be tailored to the individual experiment. For instance, it could be a routine flying task, or even an air-to-air engagement. This would be determined by the experimenter based on the type of inference that was desired from the data. Nominally, it would be anticipated that this central task would be relatively simple and essentially non-threatening to the subject. In other words, in most cases it will be desirable to obtain a pure measure of peripheral monitoring where the subject was reasonably free to carry out such monitoring. Usually, the sole function of the central task will be to assure that the subject is not able to attend to the peripheral tasks with focal attention. For this purpose, a simple flight task, such as maintaining straight and level flight in a reasonable wind condition might be appropriate. In fact, any synthetic task (e.g., a tracking task or a cognitive processing task) could also be used for this purpose.

On the other hand, in some situations, the experimenter might be interested in determining peripheral processing capability when the subject is engaged in a highly demanding central task. In this case, the present case could be combined with some of the other G-PASS tests, or with another flight simulation task which would place higher demands on the subject's focal attention. Again, the flexibility provided here yields a great deal of information which can than be used in computer simulation models.

Test 7. Unusual Attitude Recovery

General description of the test

A strong recommendation of pilot consultants on this project was that the G-PASS System incorporate some measure of a pilot's ability to quickly and appropriately recover from an unusual attitude. The operational implication of this metric is obvious. From the viewpoint of the human performance model, this task clearly involves long-term memory functions integrated with working memory to a great extent, with vestibular integrity perhaps required. The A-PASS battery recommended incorporation of a frequently used unusual attitude recovery testing technique, and this will be incorporated in to G-PASS with slight modifications, using pilot input to design the actual scenarios. The task involves having the display screen (e.g., HUD) appear suddenly in a position indicating that the aircraft is in an unusual attitude. The required response will be to

recover from the unusual attitude as quickly as possible. Measures of speed and appropriate response will be collected.

Because of the availability of a good aero-model, it is recommended that the F-16 HUD be chosen as the basic display for this test. However, virtually any aircraft for which a good model exists could be substituted. The individual will be presented with the essential information from the HUD indicating that the aircraft had assumed an attitude which, 1) could not have been anticipated by the pilot, 2) represented a dangerous condition of the aircraft. In the present case, in order to simplify the training requirements, the HUD should only present information necessary for control of the aircraft (i.e., the gun site tracking information and the ILS information will not be presented). The individual is required to take immediate and appropriate action, through use of the control inputs, to counter the unusual attitude.

Detailed test specifications

As noted above, we intend that the default stimulus display for this task will be a modified F-16 HUD display, although this could change due to operational considerations. The actual test administration can be viewed in several different contexts, determinable by the experimenter. For instance, one possible context would be that the subject would be in an isolated environment (i.e., one in which there was no previous stimulus conditions) and would suddenly see the HUD picture depicting an unusual environment. The subject's response to this condition would then be measured. In another situation, the subject might be engaged in a particular task (either a G-PASS test or a flight condition). The screen would then "blank" for some short period of time, indicating that the aircraft has entered into a "cloudbank", or other condition in which the HUD information is unavailable. At the end of some short period of time, the HUD would reappear. However, there would be an unusual attitude indicated. Again, the pilot would have to take immediate action to correct this unusual attitude in order to avoid a catastrophic outcome.

The specific unusual attitudes to be employed in the G-PASS will be determined by the experimenter. However, we will recommend at least 8 unusual attitude conditions based on the experiences of our pilot consultants. These would be roughly equated with respect to difficulty of the corrective action required.

It would be anticipated that a single trial of this test will take between 30 seconds and 2.5 minutes to complete, depending on the timing of the initial unusual attitude and the duration of the corrective action. Unlike other tests, however, it would not be anticipated that this procedure would require a large number of trials in order to arrive at a reliable estimate of the pilot's capability. In fact, two to five trials would be sufficient.

On the other hand, it should be recognized that this procedure will require some training time for non-pilot subjects, even using the simplified HUD, in order to assure that the subjects are capable of performing the corrective action required. The goal, of course, is

to assure that the subject, under nominal conditions, is perfectly able to respond to the unusual attitude in an appropriate way. On the other hand, pilot subjects should already be well trained in responding to unusual attitudes. Therefore, for these subjects, one hour of training should represent a maximum of training time. In order to assure that they are proficient, however, a sufficient number of trials should be run to guarantee stability of performance. Thus, even in pilot subjects, at least 30 minutes of training time should be considered necessary.

The subject's response to the unusual attitude will, of course, be quite idiosyncratic. Not everyone corrects the unusual attitude in the same way. Therefore, no individual measures of the *way* the subject recovers will be appropriate. Instead, the overall success or failure of the correction must constitute the basic metric of this procedure. This overall success or failure can be quantified in several ways. In the present case, the time to achieve stable flight parameters will be the basic variable. This will be assessed as a ratio relative to the time which expert pilot consider to be "optimal" for the required correction. In other words, if the expert pilots consider that the optimal correction could have been achieved in five seconds, and the subject achieved the correction in ten seconds, the ratio would be 2.0. This would then become the subject's score. Obviously, this metric is based on the assumption that the quicker the reaction (within the constraints of safety) the better it is. It will also be of interest to collect data on the subject's response time in terms of the first <u>appropriate</u> corrective action taken.

<u>Test 8. Flight Simulation Testing</u>

<u>General description of the test</u>

Although it is desirable that the G-PASS system be relatively independent of a flight simulation, it is still desirable to incorporate some reasonably high-fidelity flight simulation into the overall system. Techniques to achieve this have been studied extensively, although with limited success. Again, it is recognized that there are problems introduced when a high-fidelity aero model is introduced into a centrifuge environment. However, in view of NTI's recent experience at developing such models, and integrating them into a variety of environments, it is reasonable to attempt to incorporate such a capability into the G-PASS battery. This is probably a relatively high-risk effort, but the obvious face validity of the measures, as well as their ability to complement other measures of the G-PASS system in computer models, suggests that such an effort would be justified.

Essentially, the concept of this test procedure is to provide a realistic aero model of the F-16 aircraft that will allow the researcher to design any desired scenario or mission. The subject will "fly" that mission, either under varying G forces, or off-line after G exposure. Scoring techniques will involve standard measures of altitude, speed, and navigation errors.

TESTS ADDED TO THE G-PASS BATTERY

The above eight testing procedures are either identical to or modified from the original A-PASS battery. Given the newly developed model of human performance, however, several other testing approaches suggested themselves, and these are described below as candidates for the G-PASS battery.

Test 9. Short-term Memory with Distraction

General description of the test

Various characteristics of short-term and working memory that were mentioned above make it desirable to introduce a different kind of probe into G-PASS. This test procedure will assess three general characteristics of short-term memory. The first is simple short-term memory, ranging from a minimum of 20 seconds to a maximum of one minute. Subjects will be given instructions to carry out an activity at some point within these time limits. The second characteristic is short-term memory over a longer time period. The same procedure will be used, but the delay will range from one minute to 20 minutes. The third characteristic is short-term memory over longer periods of time with distracting tasks introduced during the interval. An additional type of probe to be introduced in this test is the nature of the task required by the subject. This will range from a simple discrete motor response, such as switch activation, to a more complex alphanumeric response at the limits of short-term storage (involving 5 to 9 chunks of information remembered). Interest will range from simply determining whether the subject remembered to carry out the action, to more complex measures of the accuracy of the response.

Detailed test specifications

The primary input modality for this test will be auditory. The subject will be told to carry out an action at some future point. The action will involve either a simple motor behavior, or a motor response based on a more complex memory and/or calculation. The actual commands to be used will be determined by the researcher as a function of the individual experiment and analysis procedures. However, we can anticipate general categories of commands, each with their own separate analysis procedures.

The first category will consist of an auditory message requesting a specific behavioral response to occur at some point 20 seconds to one minute after the command. This will require a control movement or button press. Category two will demand the same kind of motor response as above, but will be demanded between one and 10 minutes after the command is given. Category three will demand a verbal response or keyboard input, requiring the operator to remember between four and nine alpha-numeric units for one to five minutes. These can include such things as a radio frequency change, or even repeating a "code" consisting of random combinations of numbers and letters.

The researcher will be given the capability to select any number of test types for inclusion into the experiment. For each test item or instance, the researcher will have to assure that the command being given is appropriate to the scenario being flown. In other words, if any command involves an action upon reaching a certain altitude, it must be given at an appropriate time below that altitude. Radio frequency changes should be given at times that are realistic. Therefore, this task will require the researcher to identify specific initiation points appropriate for each type of question. Normally, an absolute minimum of 10 trials or incidences for each type of memory probe should be given, but it would be more desirable to present considerably more (i.e., up to 30).

Essentially, two types of response are envisioned. One would consist of discrete inputs from one or more switches, and the other would consist of a numerical input. The numerical response will be given through a keyboard, or through the radio frequency control unit. Data to be captured include the time that an action should be carried out versus the time that it actually was carried out (as well as the accuracy of the response). These will be categorized with respect to the type of memory function being tested.

Test 10. Visual Monitoring

General description of the test

In order to probe working memory functions efficiently in the context of the centrifuge, a divided attention paradigm will be used. The subject will be required to monitor systems visually while performing normal flight functions (thus, this test may be integrated with Test 8 above). The basic concept is that any of four selected display devices will indicate a degraded condition in some system. The subject must first detect the degraded conditions, and take an appropriate remedial action (switch activation or verbal response). The goal is not to make the visual detection task a threshold detection task -- the displays should be relatively easy to detect if scanned properly. This test should therefore probe the "automatic" functions of working memory.

Essentially, the test is a visual detection task, although some simple decision processes may be employed. The researcher will be permitted to introduce the detection task at any point in the mission, thereby determining the background workload level against which the task must be performed.

Detailed test specifications

The exact location and configuration of the stimulus displays for this test will depend on the final design of the centrifuge gondola. However, we recommend that the displays be placed within the subject's parafoveal vision, and that two be placed on the left extreme of that range, and two on the right. It is recommended that the displays provide numerical information (as opposed to moving dials). The dimensions of each display should approximate the actual dimensions of the displays in an F-16 aircraft in terms of visual angle subtended.

The subject's task will be to indicate detection of any numerical deviation above a pre-briefed value. Convenient, non-G-sensitive switches should be used as the response manipulanda, and should have near-millisecond accuracy. The basic data of interest in this test is time to detect a significant deviation in any display. In addition, the appropriateness or accuracy of the decision process must be captured. Therefore, separate switches should be provided that will designate which display was out of limits. The time stamp of the display entering an abnormal condition and the time stamp of the subject's responses therefore form the basic data to be collected. These response times should be labeled with respect to time into the mission, direction and type of display showing the abnormality, and the accuracy of the response.

Test 11. The Blanking Test for Assessing Situation Awareness

General description of the test

It is desirable to provide some measure of the situation awareness of the subject. Among the techniques designed to probe situation awareness, the blanking technique described by Dr. Mica Endsley is arguably the most widely used, although it is not without its critics. In this approach, the ongoing simulation is stopped or "blanked" unexpectedly, and the subject is asked a question concerning some aspect of the situation at the moment of blanking. The general concept is that if the question probes a relevant aspect of the situation at the time, the subject's answer will indicate his or her level of global situation awareness. Several kinds of questions can be asked -- at the simplest level, the question simply asks about an environmental condition at the time of the blanking (e.g., "What is your altitude?"). More complex questions might involve anticipating the actions of an enemy or friendly aircraft, or might require the pilot to manipulate two or more pieces of information in order to answer the question. The software for this test will not specify what questions should be asked in any particular application, since these will be unique to the experimental design.

Detailed test specifications

The actual presentation of the probe question is straightforward. At selected points in a scenario or mission, the screen simply blanks, and a short question is presented. The aircraft should continue to "fly" during this blanked period, so that when the display reappears, the aircraft should be in a different position than when the screen was blanked. The manual for this test will provide the experimenter with recommendations concerning where in scenarios these questions can be inserted.

In this task, the questions should be phrased in such a way that the subject's response should be constrained to "YES" or "NO", or "TRUE" or "FALSE". In other words, questions should be asked in such a way that the subject must respond with a positive or negative response. This will permit analysis of the data in terms of simple accuracy and latency measures, or in terms of a "receiver operating characteristics" (ROC) analysis.

The response manipulanda for this task should be a simple positive (yes) or negative (no) switch that is located on one of the control mechanisms. This device should have near millisecond accuracy.

Data gathered in this test will be of two types. First, the reaction time from the onset of the question to the moment of the answer will be recorded. Secondly, the correctness category of the answer will be evaluated (linked to its reaction time). The four categories of answer will be a "hit" (yes answer when the correct answer is yes), a "miss" (a no answer when the correct answer is yes), a "false positive' (a yes answer when the correct answer is no), and a "true negative" (a no answer when the correct answer is no).

Test 12. Rapid Decision Making

General description of the test

One of the most important skills considered essential to the pilot's ability is his or her decision-making capacity. It is generally believed that the ability to rapidly attend to a stimulus input, analyze it in relation to established rules and learned relationships, and then to choose between two or more alternatives through a motor action is crucial to success in the flight environment. In some cases, this obviously involves an ability to "compartmentalize" stimulus inputs so that only information relevant to the required decision is allowed to enter into it. Pilot consultants and government personnel have suggested that the radar warning receiver (RWR) display might provide an appropriate stimulus element for this type of function. In this display, a radar threat might appear at various positions on the panel. Typically, an auditory warning accompanies the appearance of the stimulus on the scope. The subject must rapidly assess the nature of the threat, as well as its severity, and decide on appropriate response. The test therefore assesses several aspects of working memory that are in the human performance model.

Detailed test specifications

The RWR simulation to be used in this test will consist of a round display containing two concentric rings creating three distinct areas within the display. The outer area will be designated as a "safe" area. No response will be required of the subject if a threat appears in this area. The middle area will be designated as a "danger" area. Depending on the nature of the target, the response of the subject to a threat in this area is either to dispense chaff and either turn away or toward the threat. The third area will be designated as a "critical" area, signifying that the threat is tracking and has fired. This type of threat will require an appropriate missile avoidance maneuver by the pilot.

Two basic kinds of threat will be used in this test. One will be an airborne threat, and the other will be a surface to air missile. These will be indicated by clearly differentiated symbology. The threats will appear in the center of four quadrants in their respective areas of the display. The position and nature of the threat will be programmable by the

experimenter. In other words, the experimenter will be able to indicate the type of threat that will occur, its frequency, it's timing, and its location. Default options will be provided. These will present each kind of threat 50 percent of the time distributed randomly and equally over the three areas and four quadrants.

The response to the appearance of an RWR threat will consist either of a chaff release, and/or an aircraft control action. The chaff release will be controlled by the normal chaff dispensing switch in the aircraft. The control actions will involve whether or not the pilot made an appropriate change in the direction of the aircraft. This will be monitored by determining the aircraft state vectors.

Data analysis will be initiated from the appearance of the threat on the RWR. This, of course, is coordinated with the auditory tone warning. The essential response criteria will be the reaction time from the start of the warning signal to the initiation of the first response (chaff or control reaction). The second criterion will be the appropriateness of the control action. The actual responses required from the subject will be determined in consultation with pilot consultants after the design of the centrifuge gondola is finalized.

DEVELOPMENT OF THE G-TOOL TO OPTIMIZE PERFORMANCE (G-TOP) AND OPERATIONAL MILITARY IMPACT (OMI) MEASURES

Introduction

One of the principal developments to be carried out under this effort involved conceptualizing a way to incorporate the assessment capability described above into the human performance model in such a way that it would be useful to the war fighter. Typically, performance models require considerable input by relatively expert users, and frequently require some degree of expert analysis of results. In the present case, the desire was to produce as close to a "turnkey" operation of the model as possible. This meant conceptualizing a vehicle in which the "expert data" would reside within the model software itself, so that the end user was required to input only a minimum of information.

In the case of acceleration stress, this means that performance results must be gathered and interpreted as part of the development program. The ultimate goal would be to develop a set of algorithms or models describing the performance effects of any G force on the human. This description would obviously involve either nominal or mean estimates of the effects, as well as their variability in a targeted population. This would be an incredibly ambitious goal for any single program. However, in the present effort, our purpose was to develop a framework that had the potential of ultimately achieving this goal. Such a framework requires the elements depicted in figure 9, and these are discussed separately below.

Database Development and Input to the Performance Model

The first requirement is to gather existing data relating G exposure to various kinds of human performance. The first and most desirable sources of input to the model will, of course, be actual performance data collected on the centrifuge or in related situations. A literature review revealed that while the existing data appear to be quite good, there is precious little of it, and virtually no standardization. It will require a major effort to consolidate and standardize these data, although some efforts in this direction have already been made (Perez, 1986).

However, it became clear that a much larger body of data exists concerning the physiological effects of acceleration stresses. Many of these studies have been summarized (Burton and Whinnery, 1996). Further, mathematical models predicting overall G-level and duration tolerances have been developed and validated using these data (Cohen, 1983; Darrah and Klein, 1986; Harding and Bomar, 1990; Burton, 1986; Burton, 2000a; Burton, 2000b). These models essentially use calculations of the physiologic and G-related interactions that result in the final level of G tolerance (usually defined as the level of G that a seated human can attain without complete loss of vision, or loss of consciousness). We believe that, since behavioral changes in terms of performance capability obviously accompany changes in brain perfusion, reliable

inferences concerning the individual's performance capability as a function of cardiac and circulatory effects of acceleration stresses can be made. While this may be difficult and arguable in many instances, such analyses will at least provide a rational basis for estimating performance effects.

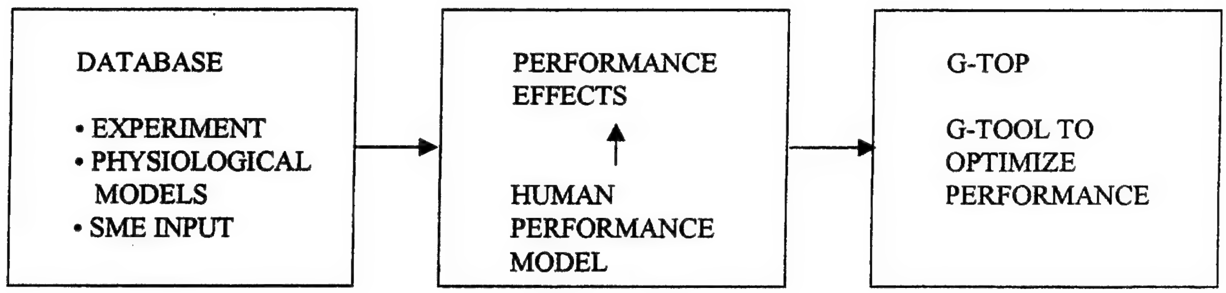

Figure 9. A framework for the development of predictions concerning the effects of G forces on performance

The third source of performance effects data can come from expert opinion. Subject matter experts (individuals who have experienced G profiles, as well as those who have studied them from a variety of viewpoints) can be used to supply reasonable estimates of performance effects for those conditions for which little data exist.

It should be noted that one of the first advantages of this approach is that unanswered research questions concerning specific types of G exposure will be clearly identified. Obviously, G effects that require the most speculation, and that are most critical to later stages of this effort will stand out. In fact, this was the major reason that the G-PASS battery described above had to be developed. Since it is designed specifically to gather data appropriate to the performance model, research using the system will input directly to the model. Therefore, any research needs that are identified can be immediately answered using the G-PASS test results.

The ideal final result of these sources of performance data would be a reasonably complete set of descriptions of the effects of G forces on each of the elements or nodes in the human performance model. In an ideal world, if empirically determined data could be entered into a valid human performance model, the final effect on operational performance of all G levels, onset types, and profiles should constitute the output. However, it is clear that this goal is unattainable in any empirical way. It simply will not be possible to investigate every conceivable condition involving G forces. Therefore, the performance model itself will have to make predictions concerning those situations that cannot be directly researched. This will require that assumptions be made concerning interactive effects, variability of effects, and other characteristics of the G experience. In addition, it will be necessary to utilize expert opinions concerning the skills required by the population of interest. These will provide the basis for integrating performance effects in each skill into an estimate of overall performance capacity.

In addressing the issue of interactive or combined effects of multiple G-exposure, for instance, it was necessary to conceptualize how we might generate predictions when data were only available for single exposures. It is likely, especially for higher G exposures, that the single curves generated from the data will not, by themselves, predict such cumulative effects. Therefore, it was necessary to incorporate a second-level "model" that attempted to describe effects of multiple G exposure. A preliminary version of this second-level model was developed, based on similar types of models in other fields, as well as on some of the physiological models noted above. Essentially, it was hypothesized that the performance decrement magnitude of a second (and subsequent) exposure to G forces is dependent on 1) the level and duration of the second exposure, 2) the physiological insult generated by the first G exposure, 3) the nature of activity intervening between exposures, 4) the type of skill being modeled, 5) the task demands on the individual during and after the second exposure, and 6) the group of factors subsumed under the heading of "individual differences".

Again, it is clear that full development of the conceptual model sketched above will require elaboration, and will still require considerable speculation and input from subject-matter-experts once it is fully developed. However, from such sources, we believe that reasonable first approximations can be developed for the initial version of G-TOP. These can then be refined as additional empirical data become available.

In summary, the task of this part of the present effort was to conceptualize a way in which a variety of data, ranging from rigorous experimental tests to subject-matter-expert opinion, could be integrated into a model of human performance. A requirement for the system is also that the output of the human performance model presents a prediction of the person's performance capacity that can then be related to <u>operational performance requirements</u>.

Development of the G-TOP Concept

Beyond the above technical requirements, the output of the human performance model, both in terms of specific and general performance capacity, must then be presented graphically to the user in a simplified and meaningful way. The goal, after all, is to permit the war fighter to evaluate operational performance capability. Therefore, a system needed to be developed that will allow the user to input unique exposure information sufficient to permit the model to estimate performance capability. Such a graphical system will actually constitute a "tool" that may be used by either the researcher or the war fighter. Our goal in this effort was therefore to develop a PC-based tool of this type, and we have designed and bread boarded the "G-Tool to Optimize Performance" (G-TOP) for this purpose.

At the time the Phase I proposal was written, we had not conceptualized how the above goal could be achieved. Only after becoming deeply immersed in the problem did it occur to us that a related effort might provide the solution. That effort was the Fatigue

Avoidance Assessment Tool (FAST) being developed by NTI as part of another Phase II SBIR effort. FAST uses a sleep model to convert specific sleep schedules into predictions of the individual's performance effectiveness. Using it, the planner can determine what work/rest schedule will yield optimal performance. With the development of a human performance model appropriate to acceleration stress, it becomes feasible to visualize a similar product. This would allow the researcher, operational planner, or even the field commander to estimate the performance effects of given schedules of G exposure, or of G exposures actually experienced by warfighters in training or combat. The above insight allowed us to attack the problems involved in filling in the details of the schematic shown in figure 9.

Prediction of single G effects

The overall concept of G-TOP is to present the "performance capacity of the individual" as a percentage of "normal" or non-degraded performance. The prototypical concept for this development is that performance prediction levels (generated either from experiment or from prediction models) under nominal 1 G conditions will be compared to those generated under various G conditions. These may have to be normalized across different skill levels to account for different units, but will result in a percentage decrement (or even improvement) estimate as a result of the G exposure (see e.g., Albery and Chelette, 1998). Where necessary because of missing data or non-standardized units, transformation algorithms may have to be used that will generate the actual performance prediction functions for each skill required by the human performance model. From these exercises, families of curves would be generated indicating expected performance change in elements of the human performance model -- sensory/perceptual functions, short-term memory, the various aspects of working memory, and response selection factors. The concept here is to estimate the impact of various levels of single G exposure on each of the skills required by the model. This will result in a family of curves such as that shown in figure 10 for a sample of the skills required by the final model described in the first section of this report. Again, since we will struggle to use existing data for all of these efforts, this constitutes a considerable literature effort.

Beyond these curves for individual nodes in the human performance model, it will be possible to construct a "composite" curve showing overall predicted decrement in performance as a result of that G exposure. In itself, this type of curve should provide general guidance to the field commander regarding the overall effect of a pilot's G history on his or her capacity to perform. However, with additional capabilities added, this can be made to be much more mission specific. For example, for a missile avoidance task it might be hypothesized that decision-making skills are relatively unimportant, while rapid reaction and perceptual recognition are critical. In the final design of the G-TOP system, extra weight could be given to those particular skills so as to tailor the "overall" performance decrement prediction to that mission.

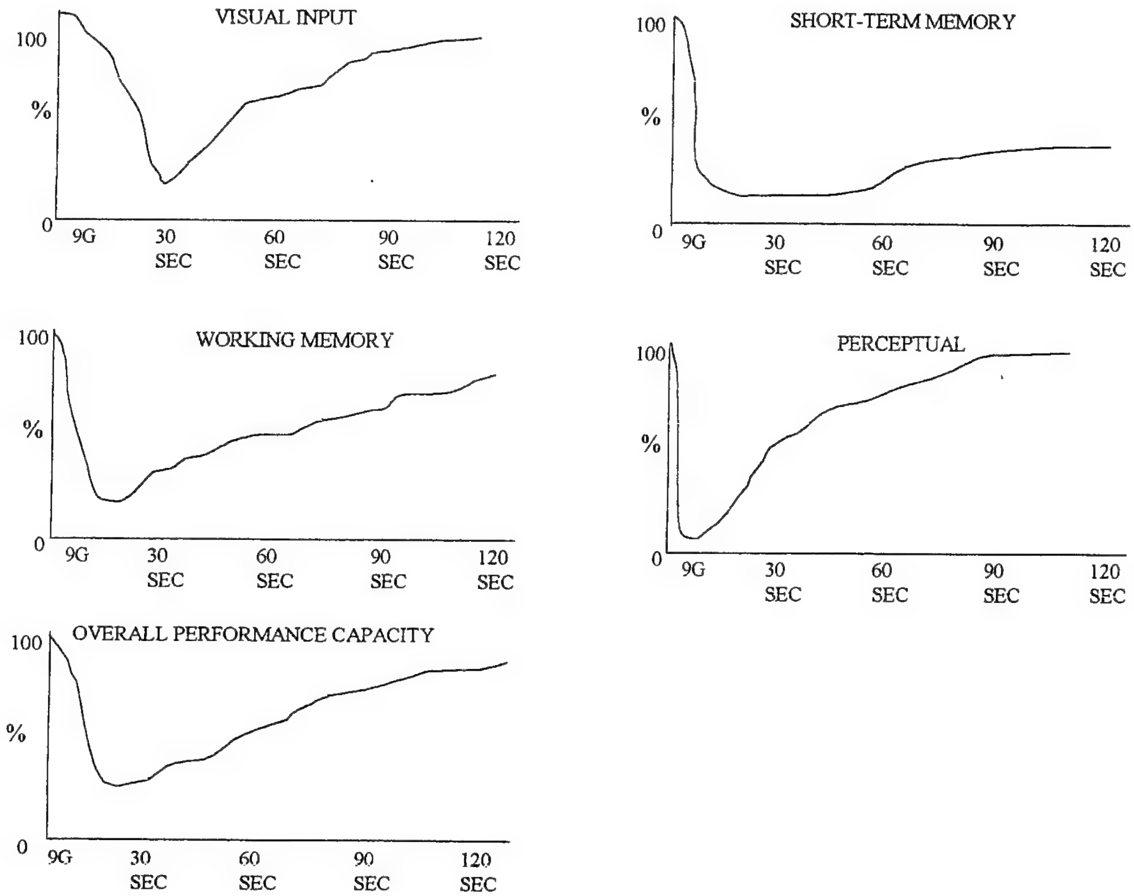

Figure 10. Hypothetical examples of performance decrements on each of the model elements from a single 9-G exposure for 30 seconds

Prediction of multiple G exposure effects

The above efforts are directed to the development of performance predictions concerning the effect of a single G exposure. These predictions constitute the underlying data set, which will give G-TOP its flexibility. The task was to conceptualize a way that combinatorial rules could be generated to integrate predications concerning any mix of G exposures. The actual details of this combinatorial approach are beyond the scope of this Phase I effort. However, initial concepts were explored, and these are illustrated in figure 11.

In this simple example, if a 9-G exposure for 15 seconds is followed 30 seconds later by a 3-G exposure, the resultant prediction for performance over the entire several minute period could be the result of the two single exposure curves additively combined. Recall that the single exposure curves are, themselves, a result of combined skill performance decrement curves determined from the performance model. The illustration shows that

the preceding 9-G exposure changes the performance prediction under 3-G in both its maximum effects, and its recovery.

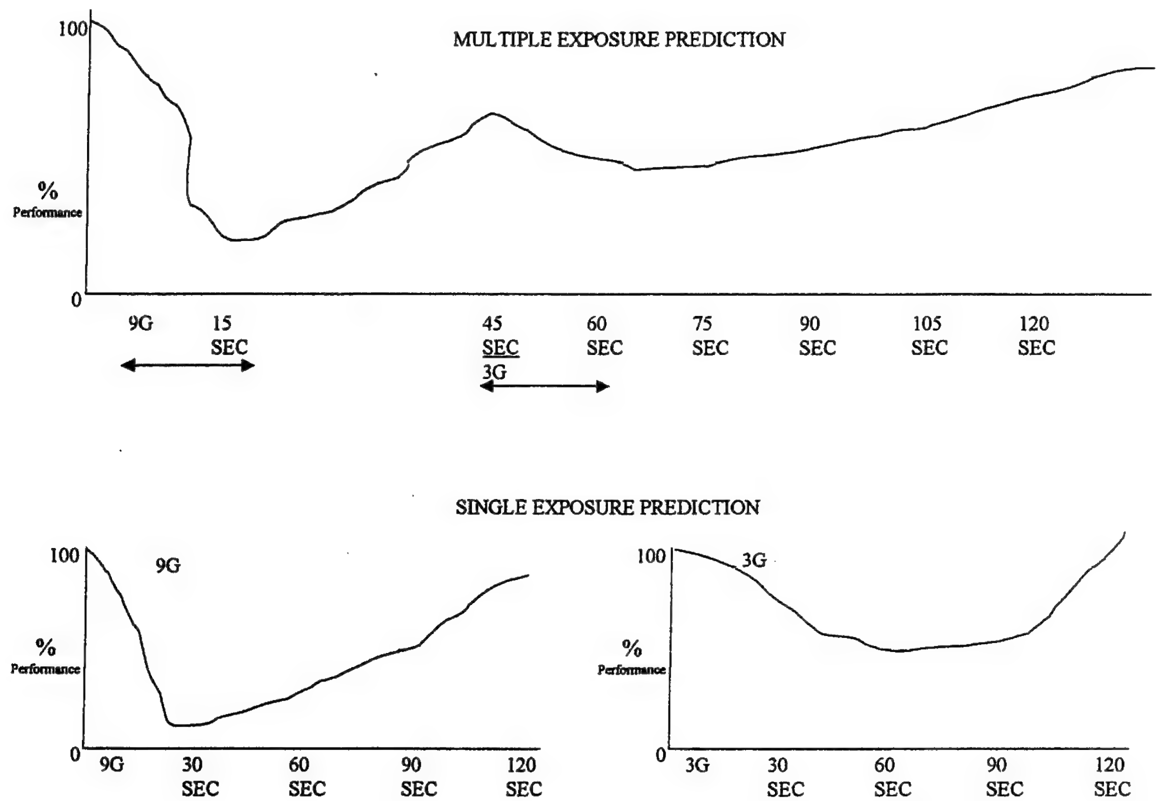

Figure 11. Hypothetical example of combined effects of two high-G exposures on performance

Of course, this is a hypothetical example since none of the actual analyses have been carried out as yet. It is possible, even likely, that existing data or models will suggest that a strictly additive approach may not always be realistic. The effect of several 9-G exposures on a subsequent 3-G exposure, for instance, may not be the same simple additive function as the effect of one prior 9-G exposure. Again, we also recognize that existing data may be sparse to generate these predictions. However, the present approach provides a heuristic framework in which assumptions are clearly labeled and the degree of certainty concerning each assumption can be at least debated.

Development of the user interface

The ultimate goal of this overall Task is to produce a user-friendly-tool that will have maximum military and other commercial application. For this reason, the technical and mathematical foundations of the tool must be presented in a most professional and ergonomically correct way. In Phase I, a demonstration of how such a tool might potentially look was provided, and this is described briefly here.

It is envisioned that the G-TOP system will consist of, at least, an animated introductory screen (figure 12a) and the performance prediction screen (figure 12b,c,d) with a series of menu selection items (figure 12e). The introductory screen will provide graphic and verbal identification of this system, as well as options concerning the next step in the procedure. Upon entering the actual system, the menus will allow the user to select specific constraints for the prediction program, and/or to further specify the conditions of the acceleration exposure. Obviously, not all of these have been defined at this point. However, some can be anticipated. For instance, one set of conditions might be envisioned for male versus female subjects, or a set of conditions or predictions might be envisioned for different types of operational missions. In addition, Configuration Wizards will be provided that lead the user through setup of common conditions.

The final screen consists of the actual data screen. We envision a presentation similar to that shown in figure 12. The user will be able to enter a G profile over a selected period of time (e.g., ranging from seconds to hours or possibly days). This will be done manually, either by using the mouse and/or, preferably, by a digital input into the keyboard. Once the desired G profile has been entered, the "enter" key will trigger the calculation of the performance prediction. The default option that will appear on the screen will likely be the prediction of overall performance capability. However, the user will have the opportunity to call up performance predictions for each of the skills involved in the underlying performance model. In other words, in addition to an overall performance capacity, specific skills such as visual motor tracking, decision making, short-term memory, etc. will be displayed using the same format as figure 12.

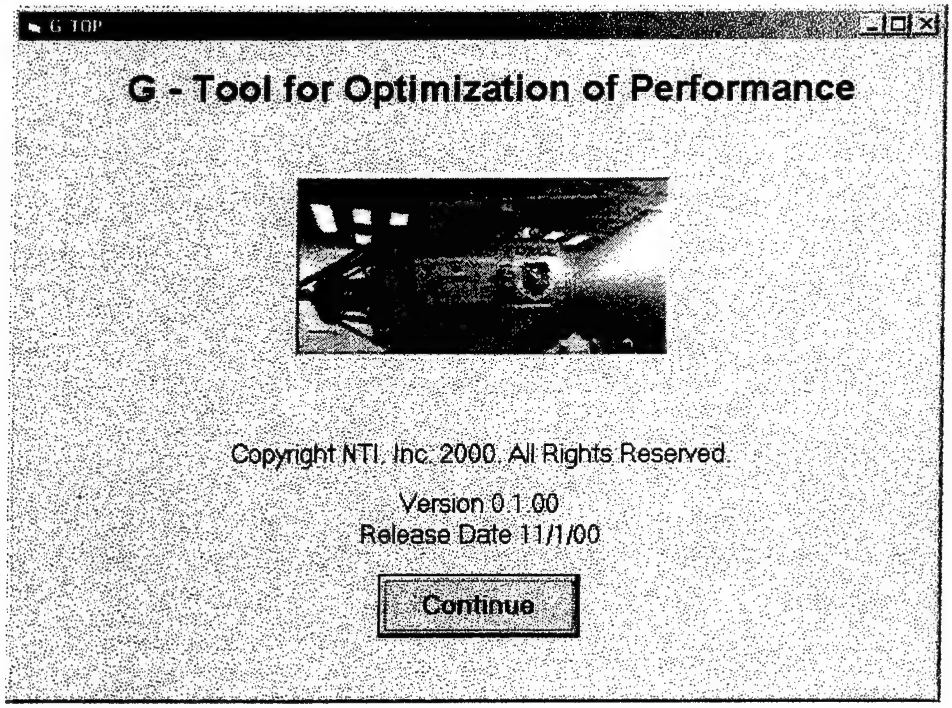

Figure 12a. Introductory G-TOP screen

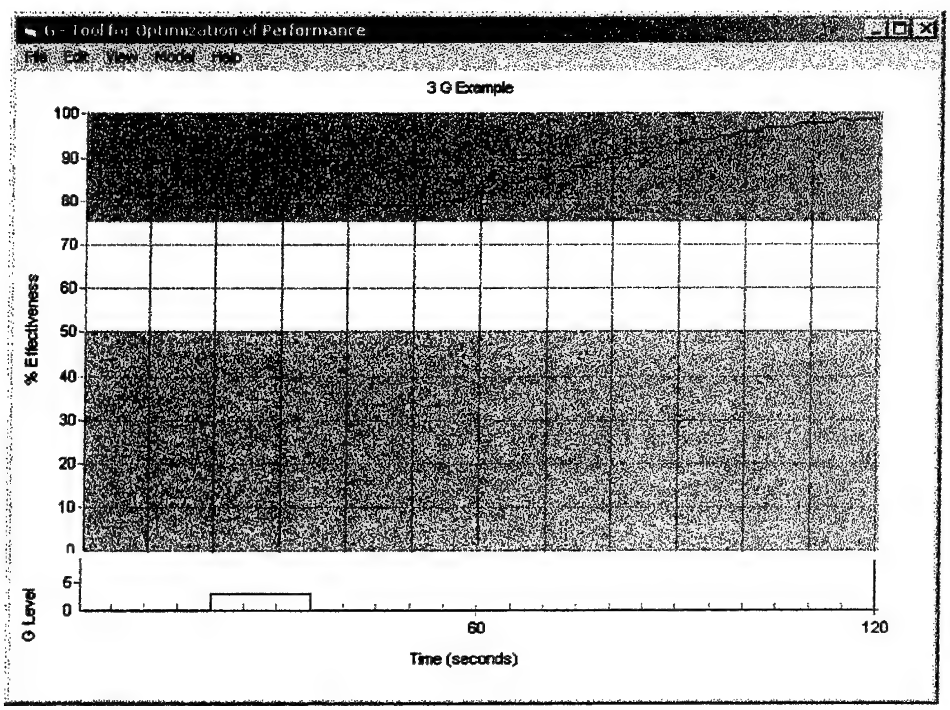

Figure 12b. G-TOP performance prediction for 3G exposure

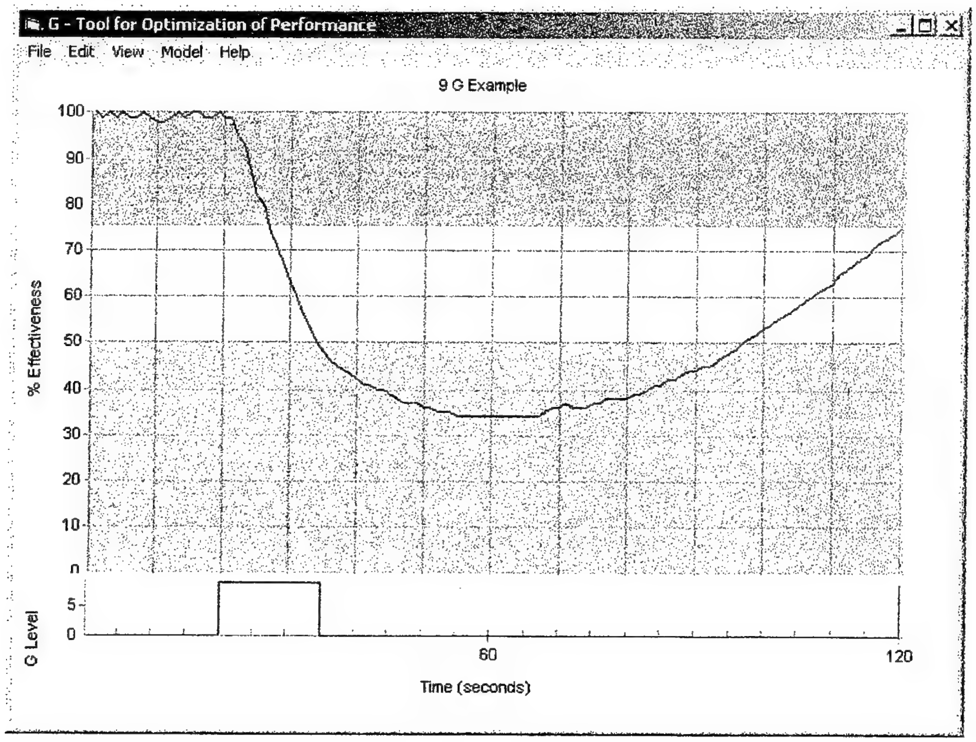

Figure 12c. G-TOP performance prediction for 9G exposure

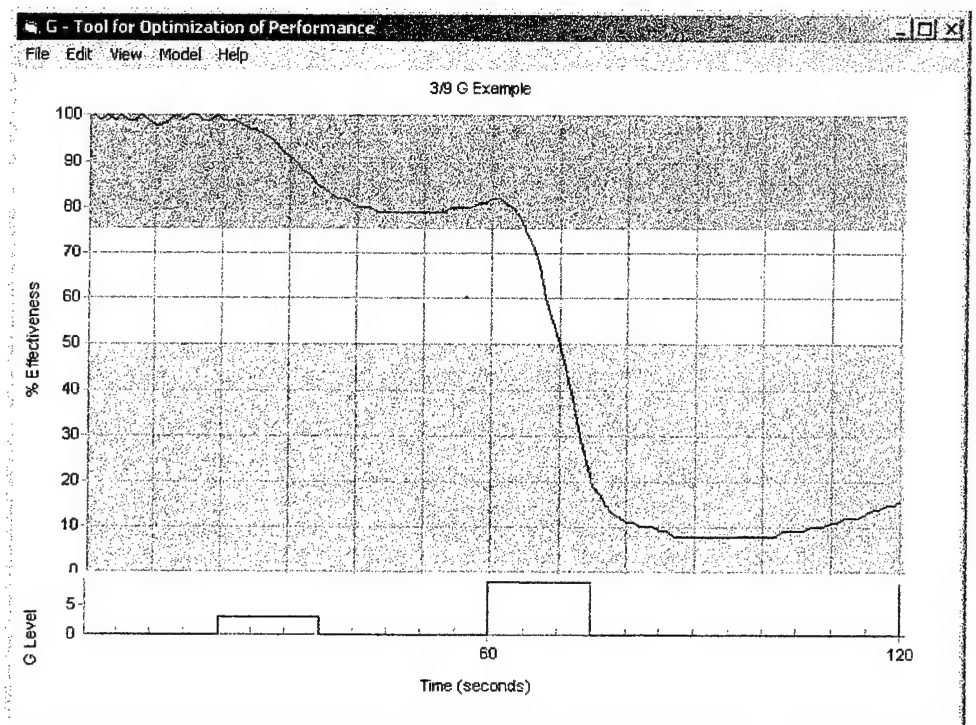

Figure 12d. G-TOP performance prediction screen for 3G followed by 9G exposure

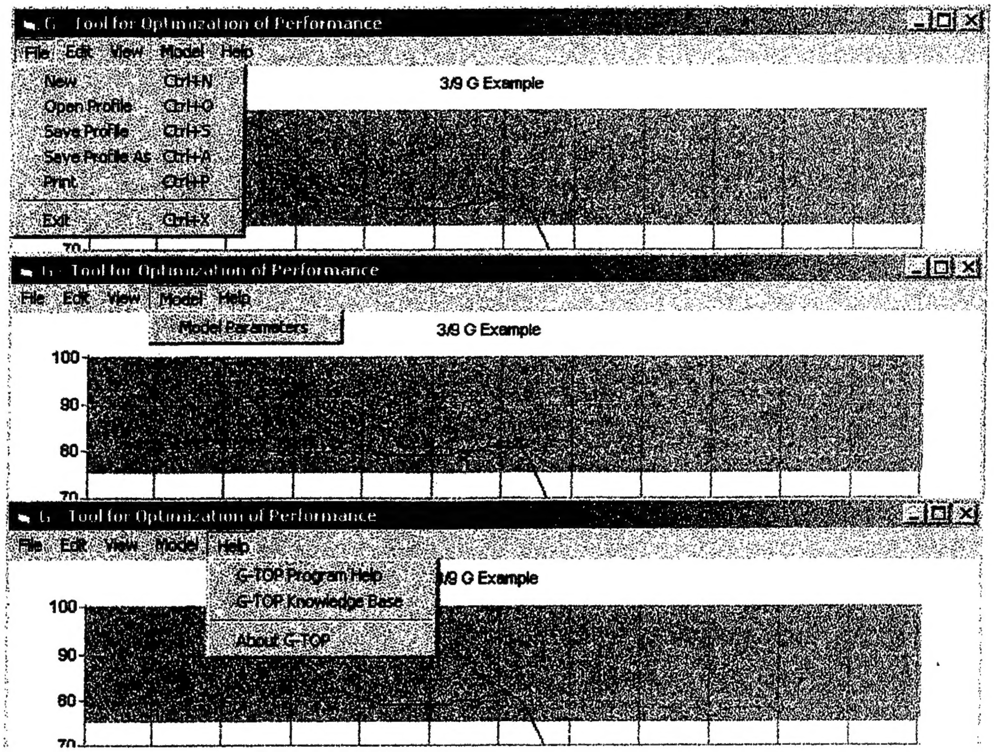

Figure 12e. G-TOP menus

Development of "Operational Military Impact" (OMI) estimates

Once performance estimates of capacity have been obtained, it will be possible to carry the process one step further. "Systems models" have been developed for a variety of aircraft applications. These range from complex multi-aircraft interactive models to high fidelity aerodynamic models of specific aircraft. For the most part, these models assume a "perfect" pilot, since they were developed primarily for engineering or operational planning purposes. However, with some modifications, some of these models can be adapted to accepting human performance data. Our concept in Phase I, therefore, was to identify how the output of the G-TOP system could be used in such a model (either already existing or to be created) to estimate the actual aircraft performance effects of human performance capacity changes. In other words, using the output of G-TOP, specific missions could be iterated many thousands of times. The success or failure of the mission as a function of G experience (the **"Operational Military Impact" - OMI**) of the G exposure could be estimated.

Specifically, we believe it will be possible to take any selected mission of interest to the Air Force, and to model that mission in software, either using existing Government models or an NTI-proprietary model (the Flight-Performance Assessment Simulation System – F-PASS) model. "Nominal" pilot reactions at each point in the missions requiring control inputs will be generated using SME opinions regarding "perfect" and "realistic" responses on the part of the pilot. These data sets will be iterated in the model

at least 1000 times each to produce a measure of merit appropriate to the mission chosen (e.g., CEP in a ground attack scenario), without the acceleration stress. A realistic G profile that would be likely to occur during or before the selected mission would then be generated in consultation with government representatives and subject matter experts. This G profile would then be run through the G-TOP system in order to determine whether any performance decrements (or enhancements) might be expected. Whatever the outcome, the percent predicted performance capability will then be entered into the mission model wherever a pilot input is required (as suggested by the human performance model). Iterating this data set the same number of times could produce an estimate of the actual OMI that would be seen in that mission subsequent to that G profile exposure.

It is recognized, of course, that the entire system above, while being based on reasonable construct-valid approaches, should ultimately be validated with appropriate criteria. As with virtually any model of extreme environments or situations (e.g., battle outcomes) criterion validation is extremely difficult and in some cases impossible. In the present case, criterion validation will eventually be possible with actual experiments carried out with live subjects on the centrifuge.

REFERENCES

Albery, W. B., and Chelette, T. L. (1998). Effect of G suit type on cognitive performance. *Aviation, Space, and Environmental Medicine*, 69(5), 474-479.

Anderson, J R. (1993). Rules of the Mind. Lawrence Erlbaum Press:Hilsdale, NJ.

Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press: Cambridge, England.

Broach, D., Goldbach, K., Weltin, M., Rosenthal, D., O'Donnell, R., and Brecht-Clark, J., (1993). Validation of the air traffic control specialist Pre-Training Screen. Paper presented at the 7th International Symposium on Aviation Psychology, Columbus, Ohio.

Broadbent, D. E. (1958). Perception and Communication. Pergamon Press: London and New York.

Burton, R. R. (1986). A conceptual model for predicting pilot group G tolerance for tactical fighter aircraft. *Aviation, Space, and Environmental Medicine,* 57, 733-744.

Burton, R. R. (2000a). Mathematical models for predicting G-level tolerances. *Aviation, Space, and Environmental Medicine*, 71, 506-513.

Burton, R. R. (2000b). Mathematical models for predicting G-duration tolerances. *Aviation, Space, and Environmental Medicine*, 71, 981-990.

Burton, R. R. and Whinnery, J. E. (1996). Biodynamics: Sustained acceleration. In R. L. DeHart (Ed.) Fundamentals of Aerospace Medicine (2nd Ed.), Williams and Wilkins: Baltimore.

Cohen, M. M. (1983). Combining techniques to enhance protection against high sustained accelerative forces. *Aviation, Space, and Environmental Medicine*, 54, 338-342.

Darrah, M. I. and Klein, E. A. (1986). Simulation of a highly dynamic G-time profile: a predictive algorithm for crewmember acceleration tolerance. In SAFE Symposium Proceedings. SAFE: Nashville.

Harding, R. M. and Bomar, J. B. (1990). Positive pressure breathing for acceleration protection and its role in prevention of inflight G-induced loss of consciousness. *Aviation, Space, and Environmental Medicine*, 61, 845-849.

Kyllonen, P. C. and Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389 – 433.

McCloskey, K. A., Tripp, L. D., Chelette, T. L., and Popper, S. E. (1992). *Human Factors*, 34(4), 409-428.

Newell, A. (1990). Unified Theories of Cognition. Harvard University Press:Cambridge, Mass.

O'Donnell, R. D., Cardenas, R., Eddy, D., and Shaw, R. L. (1995). Assessing the performance impact of G-forces: Design of the Acceleration-Performance Assessment Simulation System (A-PASS). AL/CF-TR-1996-0093. ADA320232.

Perez, W. A. (1986). Performance metrics for use under sustained Gz acceleration: Review and recommendations. (Biodynamic Data Bank Accession No. 45997). Wright-Patterson AFB: OH. Armstrong Aerospace Medical Research Laboratory.

Perez, W. A., Masline, P. J., Ramsey, E. G., & Urban, K. E., 1987. Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology. AAMRL-TR-87-007.

Shallice, T. (1988). Information-processing models of consciousness: Possibilities and problems. In A. J. Marcel and E. Bisiach (Eds.), Consciousness in Contemporary Science (pp. 305-333). Clarendon Press: Oxford.

Von Gierke, H. E., McCloskey, K., and Albery, W. B. (1991). Military performance in sustained acceleration and vibration environments. In R. Gal and A. D. Mangelsdorff (Eds.) Handbook of Military Psychology. John Wiley and Sons: New York.